Vlad-Zeno Millea

Gestionarea și analiza datelor cu *SPSS* și *PSPP*

Partea I

Introducere în gestionarea datelor statistice cu *SPSS* și *PSPP*, meniurile *File, Edit, View, Data* și *Transform*

Presa Universitară Clujeană

VLAD-ZENO MILLEA

GESTIONAREA ȘI ANALIZA DATELOR CU *SPSS* ȘI *PSPP*

PARTEA I

INTRODUCERE ÎN GESTIONAREA DATELOR STATISTICE CU *SPSS* ȘI *PSPP*, MENIURILE *FILE*, *EDIT*, *VIEW*, *DATA* ȘI *TRANSFORM*

PRESA UNIVERSITARĂ CLUJEANĂ

2018

Referenţi ştiinţifici:

Prof. univ. dr. Mihai Pascaru

Conf. univ. dr. Lucian Marina

ISBN general 978-606-37-0428-4
ISBN Vol. I 978-606-37-0429-1

Universitatea Babeş-Bolyai
Presa Universitară Clujeană
Director: Codruţa Săcelean
Str. Hasdeu nr.51
400371 Cluj-Napoca, România
Tel./fax: (+40)-264-597.401
E-mail: editura@editura.ubbcluj.ro
http://www.editura.ubbcluj.ro/

GESTIONAREA ȘI ANALIZA DATELOR CU SPSS ȘI PSPP

PRIMA PARTE
Introducere în gestionarea datelor statistice cu SPSS și PSPP, Meniurile File, Edit, View, Data și Transform

Lista graficelor

Bibliografie

1 IBM SPSS Statistics 20 Core System, User's Guide, 2011
https://www.csun.edu/sites/default/files/statistics20-core-system-guide-64bit.pdf

2 PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1
https://www.gnu.org/software/pspp/manual/pspp.pdf

3 IBM Knowledge Center, Home, SPSS Statistics 24.0.0 ONLINE
www.ibm.com/support/knowledgecenter/search/transform compute?scope=SSLVMB_24.0.0

4 SPSS Tutorials, SPSS Beginners Tutorials ONLINE
https://www.spss-tutorials.com/basics/

5 Basic Stats in PSPP For NICAR 2016 Denver / Norm Lewis, University of Florida
http://www.bikeprof.com/uploads/9/0/6/5/9065192/stats_intro_in_pspp_handout_npl.pdf

6 SPSS TUTORIAL & EXERCISE BOOK FOR BUSINESS STATISTICS, UNIVERSITY OF MISKOLC, Faculty of Economics, Institute of Business Information and Methods, Department of Business Statistics and Economic Forecasting, PETRA PETROVICS, MISKOLC, 2012
http://gtk.uni-miskolc.hu/files/10324/SPSS+Tutorial+and+excersise+book.pdf

7 DISCOVERING STATISTICS USING IBM SPSS STATISTICS, ANDY FIELD, SAGE Publications Ltd, © Andy Field 2013
http://www.fb4all.com/download/ebooks/statistics/%23Discovering%20Statistics%20Using%20SPSS%202013.pdf

8 Using SPSS For Windows Data Analysis and Graph, Susan B. Gerber, Kristin Voelkl Finn, hics, Second Edition, 2006
http://www.biometrica.tomsk.ru/lib/Susan.pdf

9 SPSS SURVIVAL MANUAL A step by step guide to data analysis using SPSS for Windows (Version 12), JULIE PALLANT, 2005
http://www.academia.dk/BiologiskAntropologi/Epidemiologi/PDF/SPSS_Survival_Manual_Ver12.pdf

10 A Handbook of Statistical Analyses using SPSS, Sabine Landau and Brian S. Everitt, © 2004 by Chapman & Hall/CRC Press LLC
http://www.academia.dk/BiologiskAntropologi/Epidemiologi/PDF/SPSS_Statistical_Analyses_using_SPSS.pdf

11 SPSS Manual for Introductory Applied Statistics: A Variable Approach, John Gabrosek, Department of Statistics, Grand Valley State University Allendale, MI USA, August 2013
http://www.gvsu.edu/cms4/asset/4CE39E3F-BF40-0D23-0D675539A8F525E7/spss.pdf

12 Official IBM SPSS Statistics 21 Brief Guide
http://www.sussex.ac.uk/its/pdfs/SPSS_Brief_Guide_21.pdf

13 SPSS ® Interactive Graphics 10.0, Marketing Department SPSS Inc. Chicago, Copyright © 1999 by SPSS Inc. All rights reserved. Printed in USA
https://sites.hks.harvard.edu/fs/pnorris/Acrobat/SPSS%20Interactive%20Graphics%2010.0.pdf

14 Using SPSS 20, Handout 3: Producing graphs ONLINE
https://docplayer.net/23500996-Using-spss-20-handout-3-producing-graphs.html

15 The Basics: SPSS for Windows, Copyright © 2002 by SPSS Inc. Chicago, All rights reserved, Printed in USA.
www.id.uzh.ch/cl/dl/sw/statmath/spss/spsskurs/spss_intro.pdf

16 IBM SPSS Statistics for Windows 22.0: A Basic, Tutorial, Social Science Research and Instructional Center ONLINE si pe capitole
https://ssric.org/node/459

17 IBM SPSS Statistics for Windows, Version 23: A Basic Tutorial, Science Research and Instructional Center ONLINE
https://ssric.org/node/582

18 Introduction to SPSS 19.0, Analytics Group, Authors: Nicholas Fritsche, Rasmus Porsgaard, Casper Voigt Rasmussen, Martin Klint Hansen, Morten Christoffersen, Ulrick Tøttrup, Niels Yding, Sørensen, Morten Mondrup Andreassen, Jesper Pedersen, Rasmus Maarbjerg, Last updated: June 2012
http://studerende.au.dk/fileadmin/www.asb.dk/servicekatalog/IT/Analysevaerktoejer/SPSS/Intro_to_SPSS_19.pdf

19 Statistică aplicată în științele socioumane. Noțiuni de bază -Statistici univariate, Cristian Opariuc-Dan Constanţa, Constanţa, decembrie 2009
https://www.researchgate.net/profile/Cristian_Opariuc-Dan/publication/215691876_Applied_statistics_in_Socio-human_sciences_-_Beginnings_Univariate_statistics/links/0fcfd5006d9921d6f1000000/Applied-statistics-in-Socio-human-sciences-Beginnings-Univariate-statistics.pdf

19 Statistica aplicata in stiintele socioumane Analiza asocierilor şi a diferenţelor statistice BIS
https://www.researchgate.net/profile/Cristian_Opariuc-Dan/publication/215691877_Applied_statistics_in_Socio-human_sciences_-_Relations_analysis_and_statistical_differences/links/0912f5006d874e3278000000/Applied-statistics-in-Socio-human-sciences-Relations-analysis-and-statistical-differences.pdf

20 Analysis of Multiple-Response Data, Petr Vlach, Miroslav Plaši, University of Economics Prague, Department of Statistics and Probability
https://statistika.vse.cz/konference/amse/PDF/Plasil+Vlach.pdf

21 SPSS for Windows Version 19.0: A Basic Tutorial
https://ssric.org/files/spss_v19.pdf

22 SPSS Step-by-Step Tutorial - Part 1
http://www.datastep.com/SPSSTutorial_1.pdf

23 Introduction to SPSS (version 23) for Windows, Practical workbook, University of Bristol
http://www.bristol.ac.uk/is/media/training/documentation/spss23-1/spss23-1t.pdf

24 IBM SPSS Statistics 23 Brief Guide
ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/23.0/en/client/Manuals/IBM_SPSS_Statistics_Brief_Guide.pdf

25 SPSS 12.0 Brief Guide, Copyright © 2003 by SPSS Inc. All rights reserved. Printed in the United States of America
https://sites.hks.harvard.edu/fs/pnorris/Acrobat/SPSS%20Brief%20Guide%2012.0.pdf

26 Using SPSS to Understand Research and Data Analysis Arkkelin, Daniel, (2014). *Psychology Curricular Materials.* Book 1.
https://scholar.valpo.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1000&context=psych_oer

27 SPSS for Beginners Copyright 1999 Vijay Gupta Published by VJBooks Inc.
http://dea128fc.free.fr/CoursB/B2-Quanti/SPSS%20for%20Beginners.doc

28 Analysing data using SPSS (A practical guide for those unfortunate enough to have to actually do it.) Andrew Garth, Sheffield Hallam University, 2008
https://students.shu.ac.uk/lits/it/documents/pdf/analysing_data_using_spss.pdf

29 Using SPSS to Understand Research and Data Analysis, Daniel Arkkelin, Psychology Curricular Materials, 2014, *Valparaiso University*
https://scholar.valpo.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1000&context=psych_oer

30 Discovering Statistics using Spss, Andy Field, 2009 SAGE Publications Ltd
http://www.fb4all.com/download/ebooks/statistics/%23Discovering%20Statistics%20Using%20SPSS%202013.pdf

31 Analysing data using SPSS (A practical guide for those unfortunate enough to have to actually do it.) Andrew Garth, Sheffield Hallam University, 2008
https://teaching.shu.ac.uk/hwb/ag/resources/material/analysingdatausingspss.pdf

32 SPSS: Data Manipulation and Advanced Topics For Windows, The Department of Statistics and Data Sciences, The University of Texas at Austin. Updated: August 2012
https://stat.utexas.edu/images/SSC/documents/SoftwareTutorials/SPSS_Advanced.pdf

33 SPSS: User Guide, Indian Council of Social Science Research, Indian Social Science Data Repository, JNU Institutional Area, Aruna Asaf Ali Marg, New Delhi - 110067 (INDIA)
http://www.icssrdataservice.in/files/ICSSR%20Data%20Service-%20SPSS%2022%20User%20Guide.pdf

34 Introduction to SPSS (version 16), for Windows Practical workbook University of Bristol Information Services document spss16-2
http://www.bristol.ac.uk/is/media/training/documentation/spss16-2/spss16-2t.pdf

35 IBM SPSS Statistics V22.0.0 documentation
https://www.ibm.com/support/knowledgecenter/en/SSLVMB_22.0.0/com.ibm.spss.statistics_22.kc.doc/pv_welcome.html

GESTIONAREA ȘI ANALIZA DATELOR CU SPSS ȘI PSPP

PRIMA PARTE

Introducere în gestionarea datelor statistice cu SPSS și PSPP, Meniurile File, Edit, View, Data și Transform

Scopul acestei cărți este de a oferi celor interesați de analiza statistică a datelor, aflați la început de drum, un suport cât mai consistent pentru inițierea în utilizarea programului SPSS (cu o largă notorietate în lumea academică) precum și a soft-ului gratuit PSPP care tinde să devină o alternativă viabilă, putând înlocui multe dintre opțiunile oferite de SPSS. Am ales prezentarea în paralel a programelor deoarece modalitățile în care cele două soft-uri tratează datele statistice sunt foarte apropiate sau identice. Evident, prezentarea în paralel a vizat procedurile din SPSS care au fost dezvoltate și în PSPP, cel de-al doilea soft oferind mult mai puține variante de gestionare a datelor statistice în comparație cu primul. Mai facem o precizare: dacă pentru SPSS există ghiduri, manuale, etc., pentru PSPP, varianta gratuită care își propune să ofere modalități similare de procesare a datelor, s-a scris foarte puțin atât în română cât și în engleză. Ca atare, în demersul nostru am pornit, de obicei, de la informațiile deținute despre funcționarea comenzilor și a diferitelor opțiuni din SPSS pe care, apoi, le-am verificat în PSPP (demers ușurat, în mare măsură, de faptul că PSPP, după cum am arătat deja, este replica gratuită a SPSS, chiar dacă e doar parțială).

Pentru verificări, exemplificări, capturi de ecran (print screen) am folosit varianta de SPSS 20 la care am avut acces la locul de muncă iar pentru programul PSPP am folosit pagina web care permite accesarea gratuită a kit-ului de instalare (http://pspp.awardspace.com/) în cea mai nouă versiune oficială a acestuia (PSPP_2017-09-09_daily_32bits) având titulatura 42.2.3.

Dorim, de asemenea, să precizăm că, pentru transparența prezentărilor noastre, am apelat, constant, la aceeași bază de date existentă pe pagina web a Fundației pentru o societate deschisă, „Baze de date de cercetare”, accesată la 1 Martie 2017 (http://datedeschise.fundatia.ro/baze-de-date-de-cercetare/), care se referă la „Implicarea civică și politică a tinerilor” (împreună cu cele două chestionare aferente: primul „val”, din 1 noiembrie 2010 și al doilea, din 2 iunie 2011).

Lucrarea pe care o propun pentru început vizează gestionarea datelor cu programele SPSS și PSPP prin intermediul primelor cinci meniuri, File, Edit, View, Data și Transform și reprezintă prima parte dintr-un ciclu de 3 lucrări pe care ne propunem să le realizăm; a doua parte ar urma să prezinte procedurile programelor SPSS și PSPP care corespund meniurilor Direct Marketing, Graphs, Utilities, Add-ons, Window și Help. Partea a treia își propune să se constituie într-o introducere atât în statistica uni și bivariată cât și în descrierea facilităților oferite de programele PSPP și SPSS pentru realizarea acestui tip de analize.

În viitor sunt interesat de realizarea unor duble prezentări, atât a unor modalități mai complexe de analiză a datelor statistice cât și de descriere cât mai clară a suportului pe care programele SPSS și PSPP le oferă in acest domeniu.

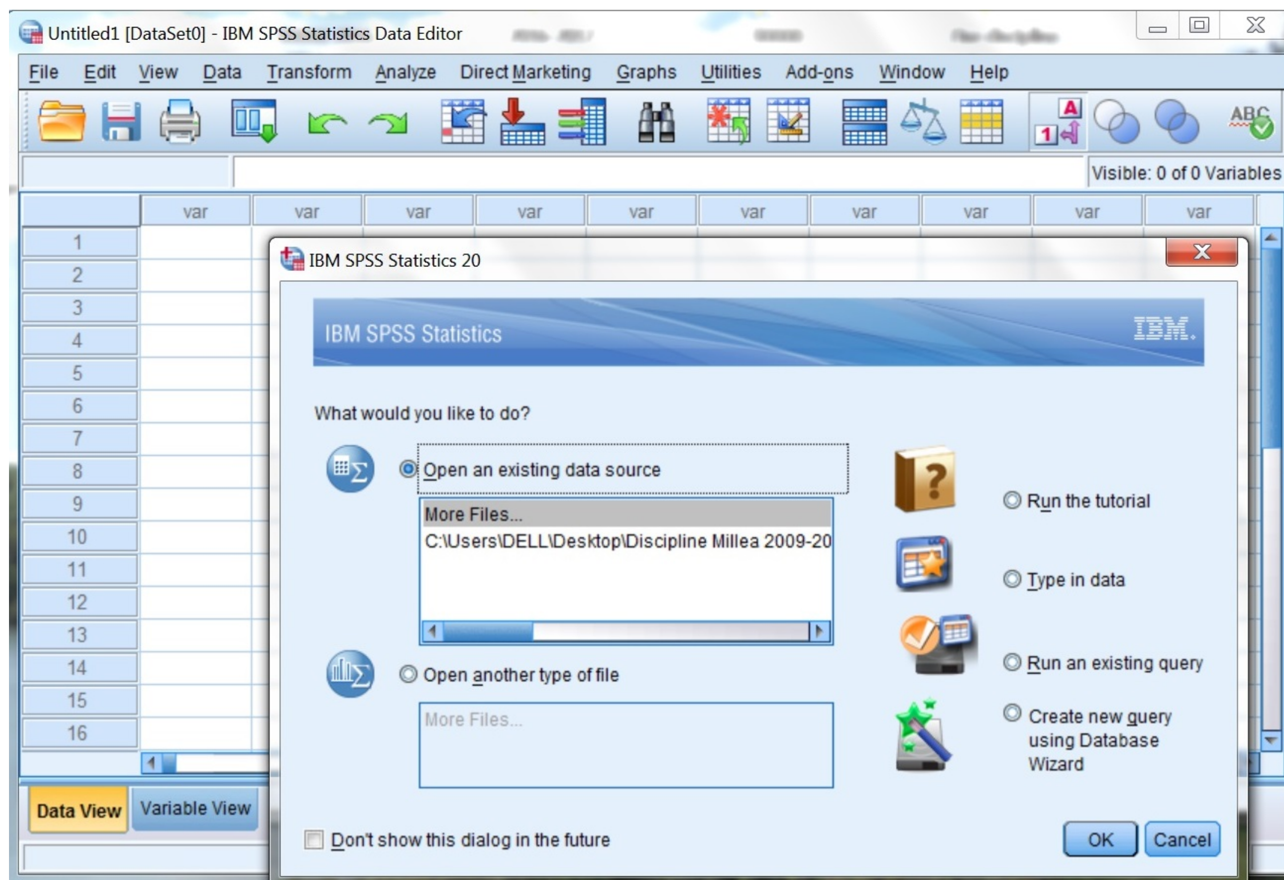
Consider că demersul nostru este util în primul rând deoarece popularizează programul PSPP prezentând situațiile în care reprezintă o alternativă viabilă pentru SPSS și pentru că meniurile programelor sunt prezentate în manieră sistematică, încercând să releve toate variantele disponibile de gestionare a datelor.

1. ASPECTE GENERALE VIZÂND PROGRAMELE SPSS ȘI PSPP

1.1 Lansarea programului SPSS și PSPP

Inițializarea SPSS se realizează ca pentru orice aplicație care rulează sub Windows, prin succesiunea **Start – Programs - IBM SPSS Statistics 20** sau dând dublu clic pe pictograma existentă pe desktopul calculatorului (IBM SPSS Statistics for Windows 22.0: A Basic, Tutorial).

Graficul 1.1 Inițializarea SPSS



După inițierea programului SPSS, este lansată fereastra (domeniul) „Data Editor”, care permite gestionarea datelor în SPSS și o fereastră de dialog, prin intermediul căreia se pot alege diferite variante de inițiere a sesiunii de lucru SPSS:

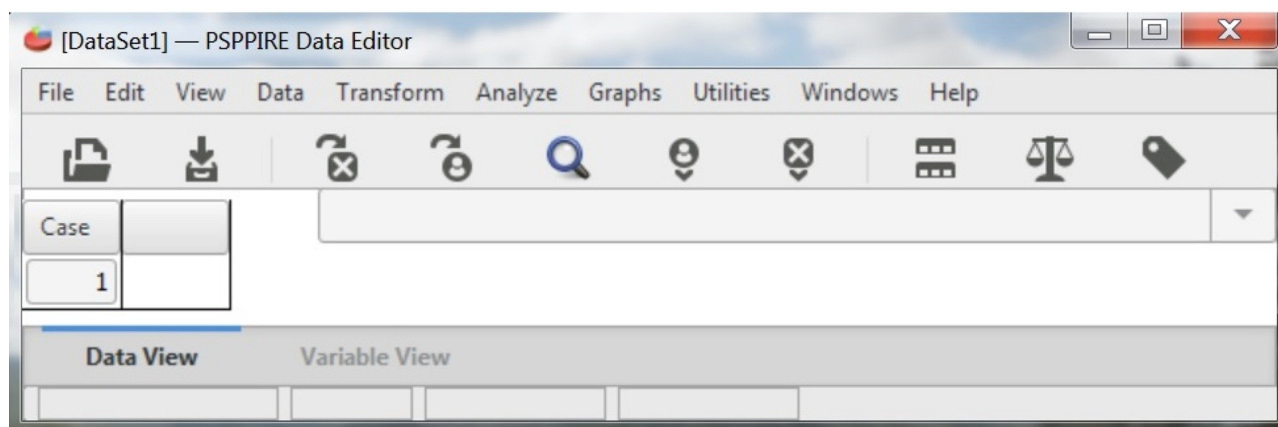
– „Run the tutorial” presupune lansarea unui text suport explicativ prrivitor la funcționarea programului SPSS
– „Type in data” desemnează crearea unei noi baze de date în SPSS
– „Run an existing query” sau „Create new query” presupun lansarea unui protocol de importare (selectivă) a datelor dintr-o bază de date externă (protocol care există deja sau urmează a fi creat)
– „Opening an existing data source” propune deschiderea unei baze de date cu care s-a operat în programul SPSS
– „Opening an other type of file” permite deschiderea oricărei baze de date SPSS sau a altui tip de fișier compatibil (creat în SPSS sau în alt program).

Pentru detalii, consultaţi IBM SPSS Statistics 20 Core System, User’s Guide.

Dacă vom da clik pe „cancel" în partea din dreapta jos a ferestrei de dialog, va rămâne activă fereastra din fundal „Data editor" (graficul 1.1), care permite atât gestionarea datelor cu SPSS cât și analiza statistică a acestora (incluzând crearea reprezentărilor grafice). În capitolele următoare ne vom ocupa, preponderent, de modul în care datele sunt evaluate, transformate și pregătite pentru analiză.

În programul PSPP lansarea e mult mai simplă și conduce la apariția unei ferestre având funcție similară cu „IBM SPSS Statistics Data editor" (a se vedea „PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1").

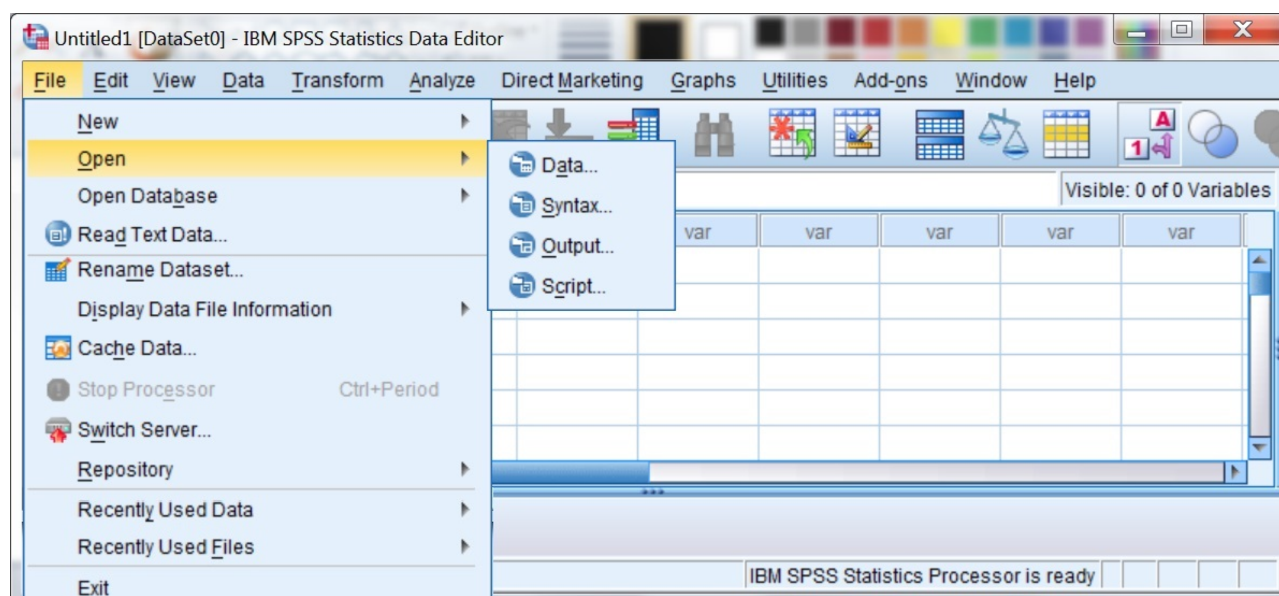
Graficul 1.2 Inițializarea PSPP



1.2 Deschiderea unui fișier în programul SPSS și PSPP

Meniul „**File**" permite deschiderea mai multor tipuri de fișiere cu care operează SPSS.

Graficul 1.3 Deschiderea unui fișier în SPSS



În pagina web „IBM Knowledge Center, Home, SPSS Statistics 24.0.0" se pot obține informații înteractiv vizînd comenzile SPSS (inclusiv „File", „Open", evident).

Folosind succesiunea de comenzi **File, Open, Data**, programul lansează o fereastră de dialog în care putem opta pentru deschiderea mai multor tipuri de fișiere tip bază de date, generate de programe cum ar fi SPSS (cu extensia .sav), EXCEL (cu extensia .xls, .xlsx sau .xlsm), dBase (cu extensia .dbf), etc.

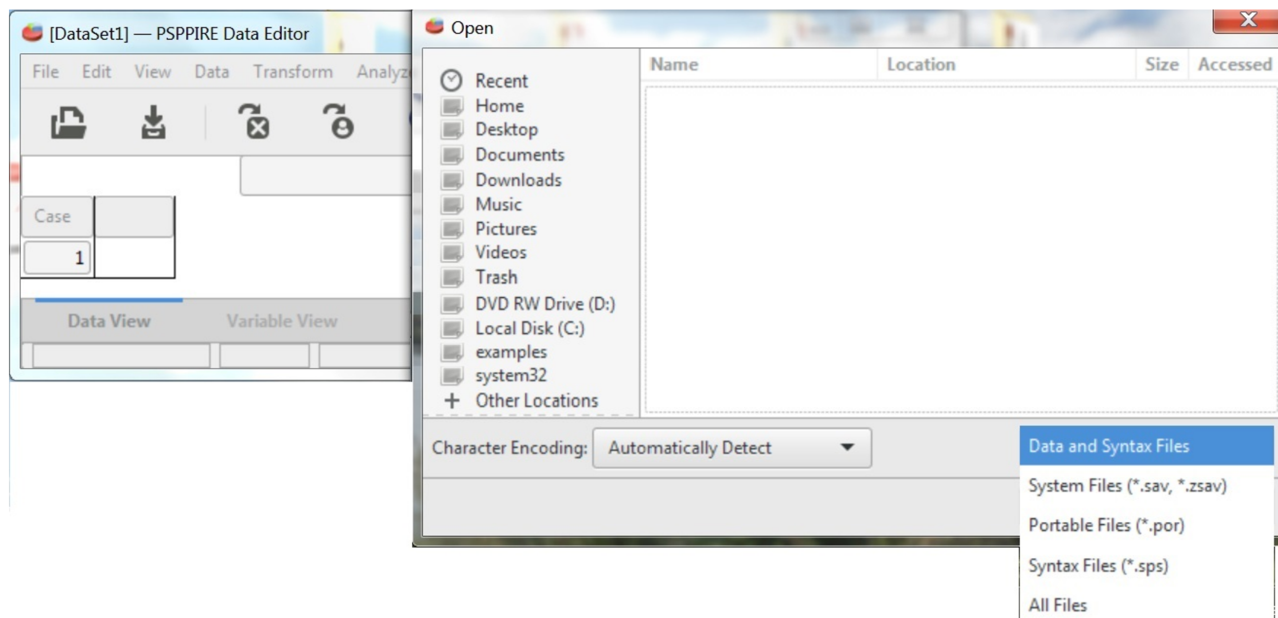
File, Open, Sintax permite deschiderea unei ferestre speciale, în care pot fi scrise comenzi ale programului SPSS folosind limbajul propriu acestuia. De asemenea, există posibilitatea înregistrării succesiunilor de comnezi date din meniurile SPSS în interiorul unei ferestre Sintax cu scopul de a fi reutilizate pentru prelucrări similare de date (având extensia .sps).

File, Open, Output dă posibilitatea deschiderii unui fișier conținând prelucrări de date sau mesaje ale programului SPSS, având extensia .spv.

File, Open, Script presupune deschiderea unor fișiere având extensia .sbs care conțin succesiuni de comenzi utile pentru eficientizarea acțiunilor repetitive din SPSS sau reiterarea unora de mare complexitate.

Folosind comanda **File, Open** în PSPP vom obține o fereastră de dialog în care, în partra stângă, sunt propuse surse ale fișierelor ce urmează a fi deschise (Desktop, Documents, Dowlnloads, etc.) Dând click pe butonul din dreapta jos al ferestrei „Open", având titulatura de „**Data and Syntax Files**", constatăm că trei tipuri de fișiere pot fi deschise (a se vedea pagina web „SPSS Tutorials, SPSS Beginners Tutorials", „Introduction to SPSS".

Graficul 1.4 Deschiderea unui fișier în PSPP

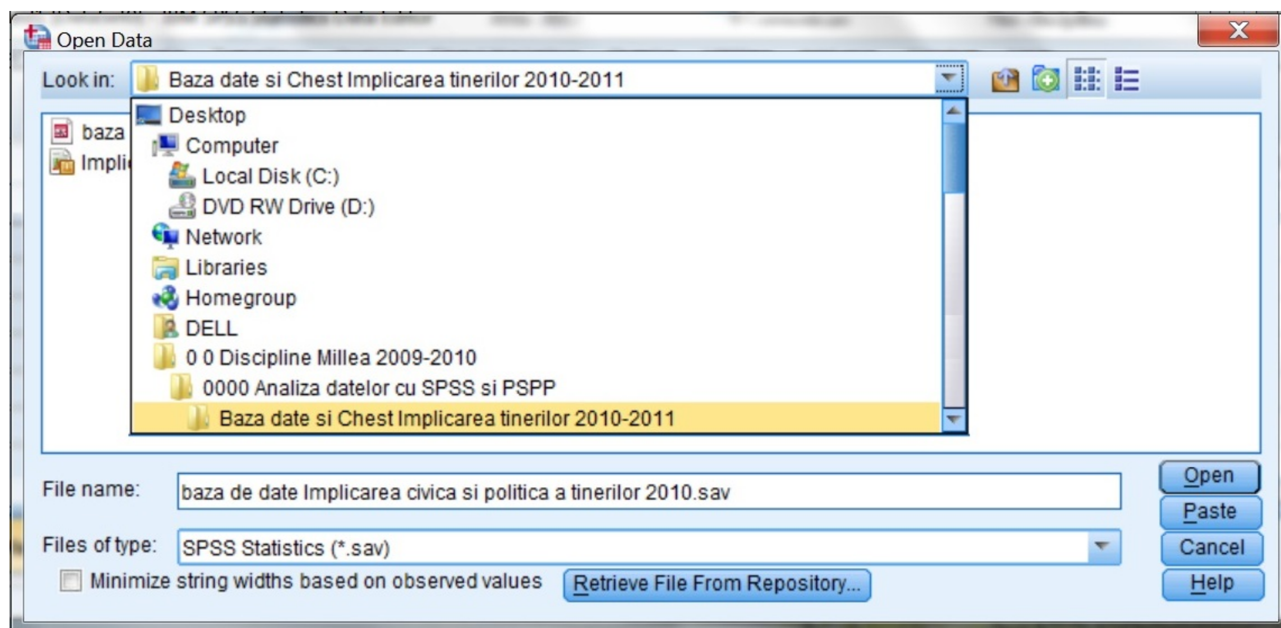


- Fișirerele tip bază de date SPSS având extensia .sav
- Fișirerele tip bază de date comprimată, cu extensia .zsav, care nu pot fi deschise cu programul SPSS
- Fișiere utilizate pentru transferarea bazelor de date SPSS (între diferite medii de programare) având extensia .por, pot fi utilizate atât în programul SPSS cât și în PSPP.

Notăm, deci, că doar fisierele de tip bază de date (incluzând variantele cu extensia .por, utile în transmiterea datelor) pot fi accesate din ambele soft-uri (fișierele cu prelucrări de date sunt specifice fiecărui program). Pentru mai multe detalii, consultați ghidul de utilizare al programului (PSPP Users' Guide GNU).

Pentru deschiderea unui fişier în SPSS (ca în orice alt soft care rulează sub Windows) e necesar să indicăm calea de acces spre acesta.

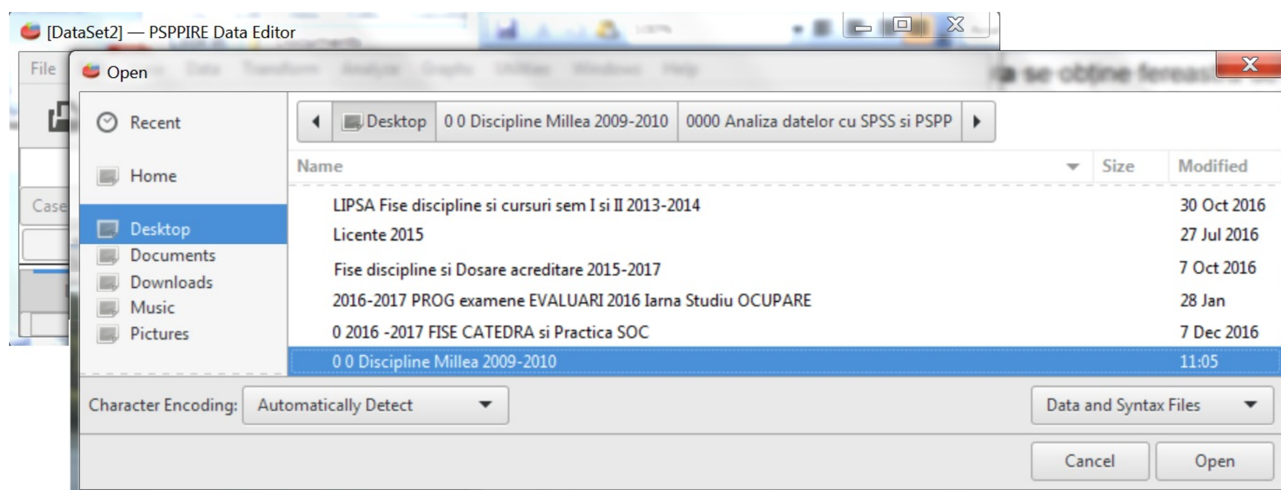
Graficul 1.5 Indicarea accesului spre fișierul ce va fi deschis în SPSS



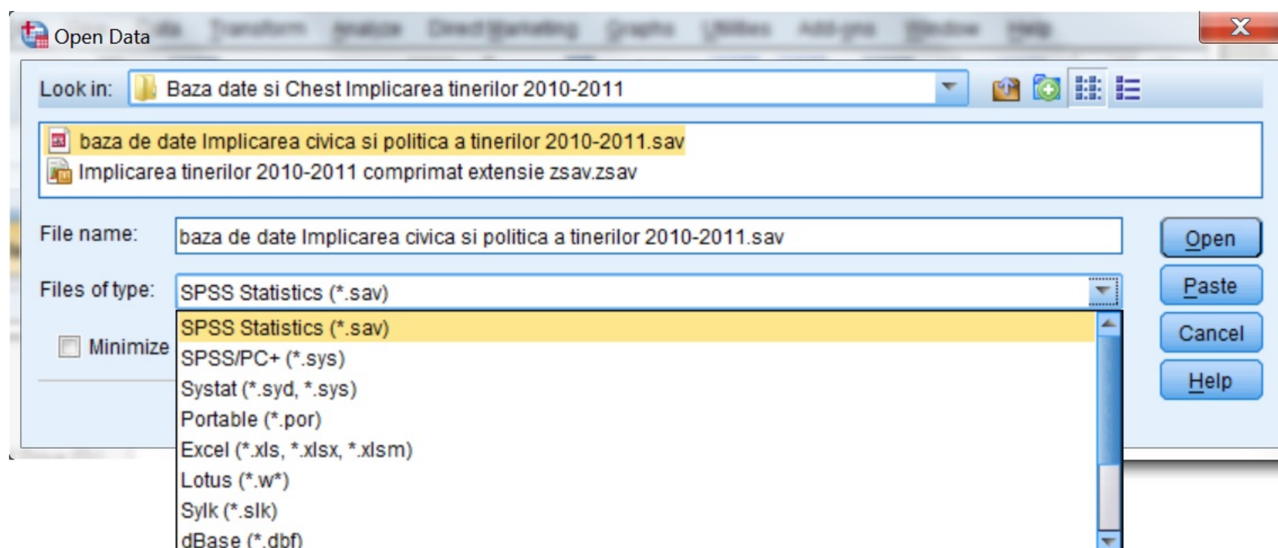
Prin succesiunea **File, Open, Data** se obține fereastra de dialog vizibilă în Graficul 1.5. Opțiunea „Look in" ne permite să indicăm modul în care poate fi accesat fișierul (de tip bază de date, în cazul nostru) pe care dorim să-l deschidem. Inițial am selectat „Desktop" după care am ales „0 0 Discipline Millea 2009-2010", apoi „0000 Analiza datelor cu SPSS si PSPP" și, în final, folderul „Baza date si Chest Implicarea tinerilor 2010-2011" în care se găsește baza de date care ne interesează (baza de date Implicarea civica si politica a tinerilor 2010.sav). Detalii în plus există în „SPSS TUTORIAL & EXERCISE BOOK FOR BUSINESS STATISTICS".

În PSPP lucrurile se desfășoară similar: în partea dreaptă a ferestrei de dialog selectăm „Desktop" apoi, din dreapta ferestrei, se alege, cu dublu click, „0 0 Discipline Millea 2009-2010" și „0000 Analiza datelor cu SPSS si PSPP" (Basic Stats in PSPP For NICAR 2016 Denver).

Graficul 1.6 Indicarea accesului spre fișierul ce urmează a fi deschis în PSPP



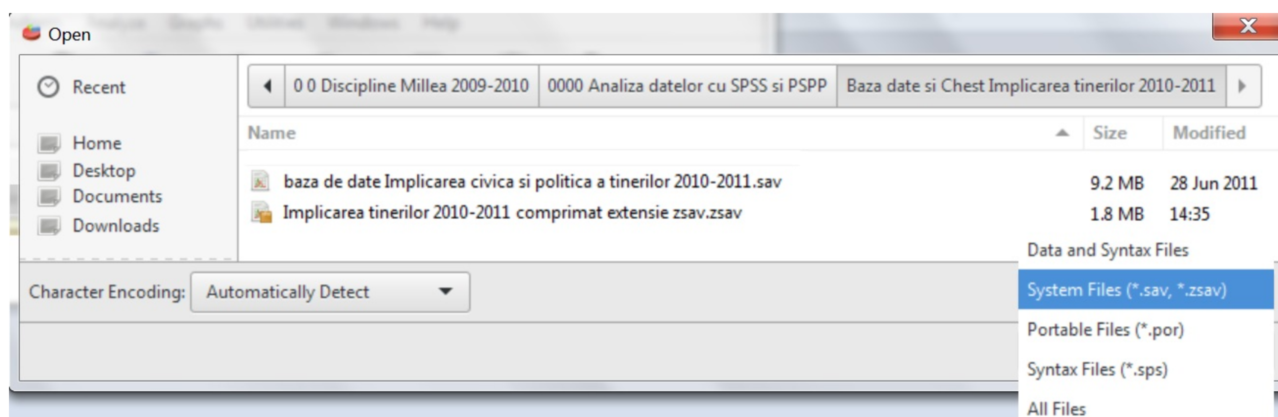
Graficul 1.7 Selctarea tipului de fișier vizualizat în fereastra de dialog SPSS



În fereastra de dialog care a permis selectarea folderului conținând informația vizată de noi există, de asemenea, posibilitatea selectării tipului de fisier care să fie vizualizat (în mod implicit, SPSS vizuaizează bazele de date cu extensia .sav). Dacă am fi selectat, în căsuța „Files of type" varianta „Excel", în partea de sus a ferestrei de dialog ar fi fost afișate toate fișierele cu extensia .xls, .xlsx respectiv .xlsm din folderul „Baza date si Chest Implicarea tinerilor 2010-2011", selectat anterior în căsuța „Look in".

În graficul 1.7 este prezentat modul de selectare a tipului de fișier care să fie vizualizat în partea centrală a ferestrei de dialog. Selecția se face de la butonul din partea stângă-jos a ferestrei, denumit „**Files of type**" (SPSS TUTORIAL & EXERCISE BOOK FOR BUSINESS STATISTICS).

Graficul 1.8 Selectarea tipului de fișier vizualizat în fereastra de dialog PSPP

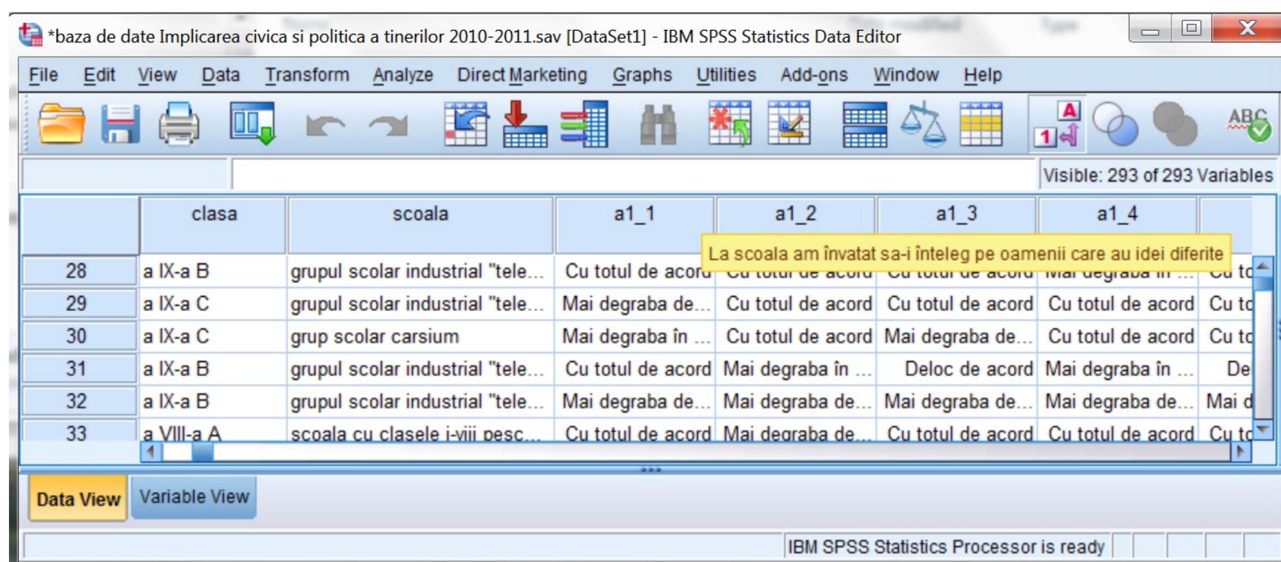


Opțiunea e similară cu ce existentă în SPSS, dar există doar variantele de fișier bază de date (cu extensia .sav respectiv .zsav pentru cele comprimate), la care se adaugă un fișier transferabil (cu extensia .por) respectiv fișierele conținâd comenzi SPSS (Syntax Files, cu extensia .sps).

PSPP este evident inferior din perspectiva utilizării bazelor de date create în alte medii de programare (nu trebuie să uităm, însă, că reprezintă, în primul rând, o alternativă gratuită la softul licențiat SPSS (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

Revenind la programul SPSS, după ce am selectat fișierul tip bază de date (cu extensia .sav), dând click pe butonul „**Open**” vom deschide o nouă fereastră care prezintă datele existente în „baza de date Implicarea civica si politica a tinerilor 2010-2011”.

Graficul 1.9 Fereastra Data Editor în SPSS (vizualizarea etichetelor valorilor)



Numele fiecărei variabile din baza de date apare în partea de sus a tabelului iar fiecare linie a aceluiași tabel reprezintă o unitate statistică (în exemplul propus de noi, elevii claselor VIII-XII cuprinși în 5 sub-eșantioane cu reprezentativitate la nivel național). Deplasând mouse-ul deasupra numelui unei variabile, putem afla care este eticheta acesteia (text care are rolul de a explicita semnificația respectivei variabile) De exemplu, variabila cu numele „a1_1” are eticheta „La scoala am învatat sa-i înteleg pe oamenii care au idei diferite”.

În general, în SPSS datele sunt introduse sub formă numerică pentru creșterea rapidității procedurii. Ca atare, e nevoie să se aloce etichete valorilor numerice pentru stabilirea semnificației acestora. Alegerea de a vizualiza etichetele sau valorile se face prin comenzile **View, Value Labels** (Discovering Statistics using IBM SPSS Statistics).

Graficul 1.10 Fișierul Data Editor în SPSS (vizualizarea valorilor)



Pentru variabila a1_1 („La scoala am învatat sa-i înteleg pe oamenii care au idei diferite"), valoarea 4 are eticheta „cu totul de acord", valoarea 3, mai degrabă de acord, 2 „mai degrabă în dezacord" iar 1 are eticheta „total dezacord". Pentru alte exemple, a se vedea „Using SPSS For Windows Data Analysis and Graph".

În PSPP există o opțiune identică cu cea din SPSS privitor la vizualizarea valorilor sau etichetelor acestora atunci când este inspectată baza de date (**View, Value labels**). De asemenea, prin poziționarea mouse-ului deasupra numelui unei variabile (din patea de sus a tabelului) va fi vizualizată eticheta respectivei variabile. Pentru variabila a1_2 e prezentată eticheta acesteia, adică „La şcoală am învăţat să lucrez în grup, împreună cu alţi elevi"; analog se procedează pentru vizualizarea celorlalte etichete.

Graficul 1.11 Fișier bază de date în PSPP (vizualizarea etichetelor și valorilor)

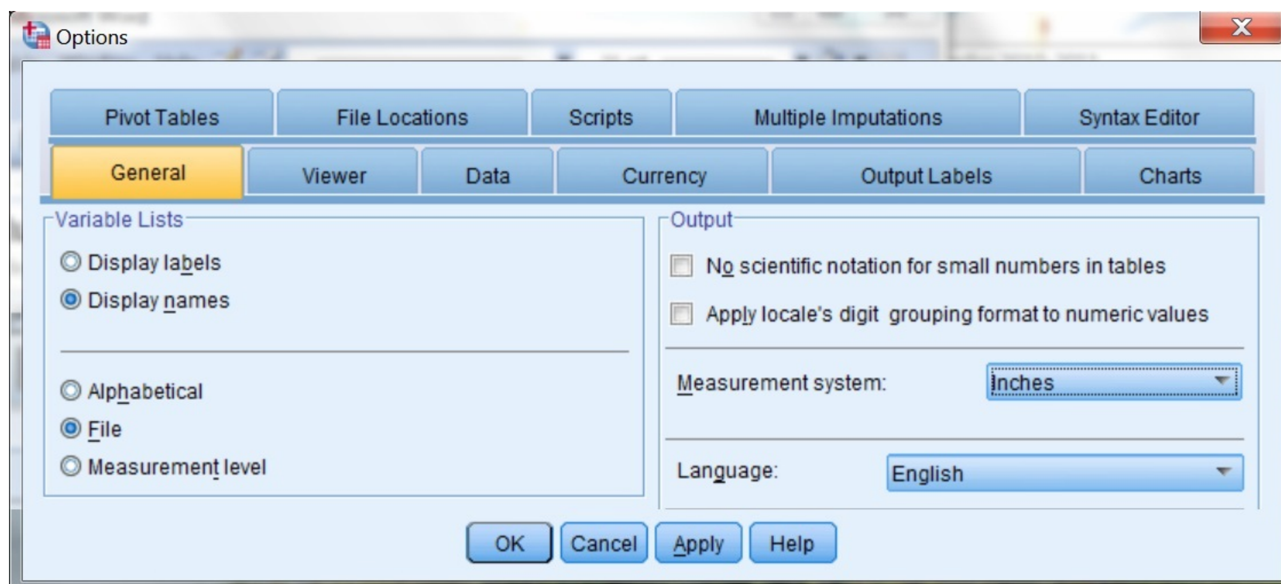
Pentru aceeași variabilă, în partea din stânga a graficului 1.11 apar etichetele valorilor iar în partea dreaptă, valorile efectiv intoduse în baza de date. De exemplu, pentru valoarea 3 a varabilei a1_1 eticheta este „Mai degrabă de acord" iar pentru valoarea 2, eticheta e „Mai degrabă în dezacord" (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1)

1.3 Lansarea unei proceduri de analiză a datelor în SPSS și PSPP

Lansarea unei analize a datelor în SPSS se realizează folosind meniul **Analyze**, care, după ce a fost deschis, oferă o largă varietate de proceduri, fiecare, la rândul ei, având, în general, mai mlte opțiuni. În această parte introductivă ne vom limita la prezentarea succintă a procedurii „**Descriptive statistics**". Înainte de a arăta cum funcționează, credem că e important să menționăm că listele cu variabile din SPSS din căsuțele de dialog ale progrmului pot apărea sub forma numelor respectivelor variabile sau al etichetelor atribuite acestora. În plus, ordinea de prezentare poate fi cea în care diferitele variabile au fost introduse în baza de date dar se poate opta și pentru vizualizarea lor în ordine alfabetică sau în funcție de nivelul de măsurare al acestora (cantitativ, calitativ ordinal respectiv nominal).

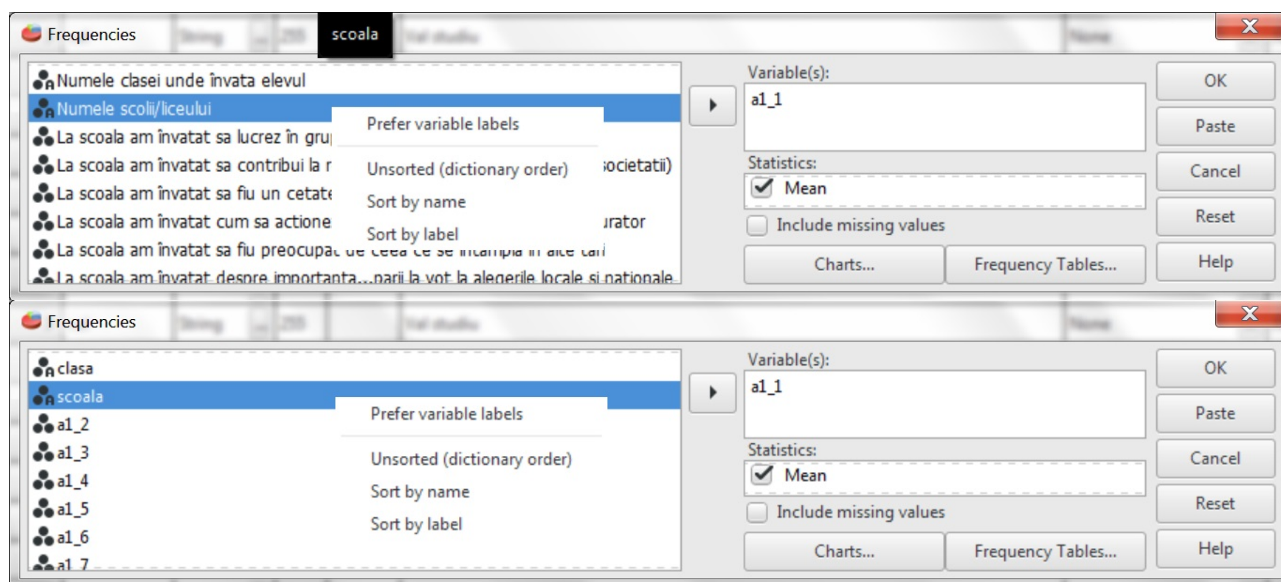
În graficul 1.12 fereastra de dialog – prezentată parțial – s-a obținut prin comenzile **Edit, Options, General**. În partea din stânga sus a feresteri se poate opta pentru vizualizarea etichetelor respectiv a numelor variabilelor iar în stânga jos putem alege ordinea de prezentare a acestora (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 1.12 Setarea prezentării variabilelor in căsuțele de dialog SPSS



În programul PSPP selecatarea vizualizării numelui / etichetei variabilei și a ordinii de prezentare a caracteristicilor din baza de date se realizeză în interiorul căsuțelor de dialog care conțin liste cu respectivele caracteristici (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

Graficul 1.13 Setarea prezentării variabilelor în casutele de dialog PSPP

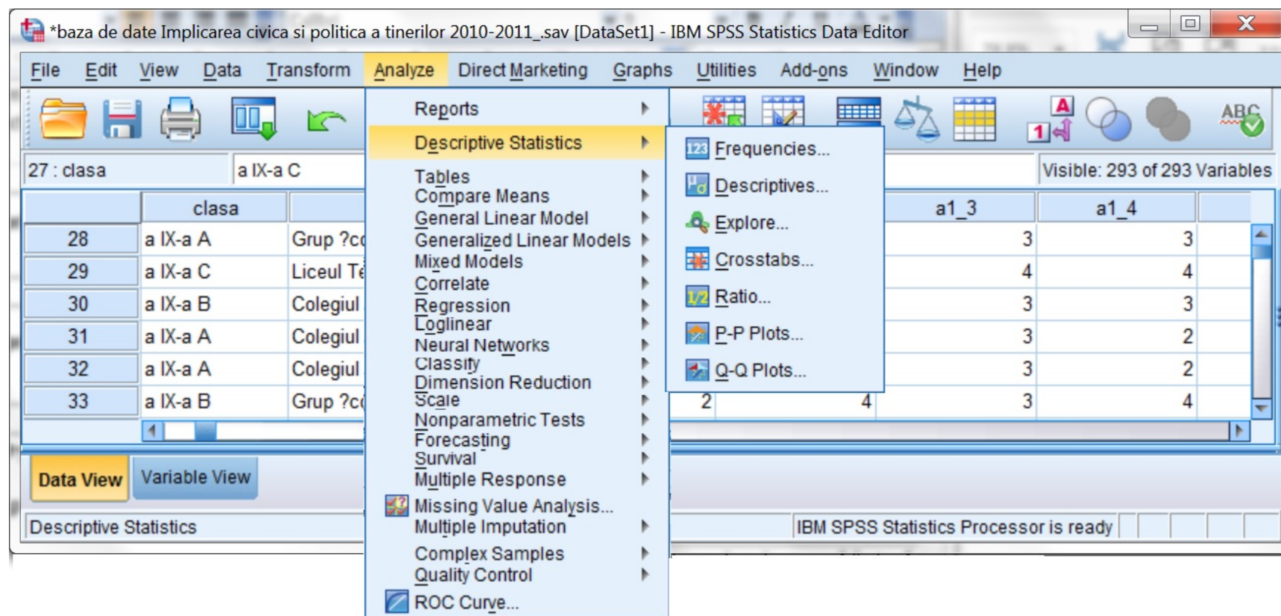


În mod implicit, în fereastra de dialog deschisă în PSPP prin comenzile **Analyze, Descriptive statistics, Frequencies** se obține varianta din partea de sus a graficului 1.13, unde sunt vizualizate etichetele variabilelor iar ordinea este cea existentă în baza de date (Unsorted – dictionary order). Dacă dăm click drepta asupra numelui unei variabile apare o căsuță suplimentară unde, în partea de sus, există opțiunea „Prefer variable labels". Alegând această opțiune vom face să se afișeze numele variabilelor dacă înainte erau afișate etichetele și, invers, vom determina afișarea etichetelor dacă anterior erau afișate numele (Basic Stats in PSPP For NICAR 2016).

În partea de jos a căsuței de dialog din PSPP putem opta pentru sortarea implicită a variabilelor (ordinea din baza de date), pentru ordonarea după numele variabilelor respectiv în funcție de etichetele acestora.

Descriem succint, pantru SPSS, calcularea distribuției de frecvență și a ponderilor valorilor variabilelor. (SPSS SURVIVAL MANUAL, A step by step guide to data analysis)

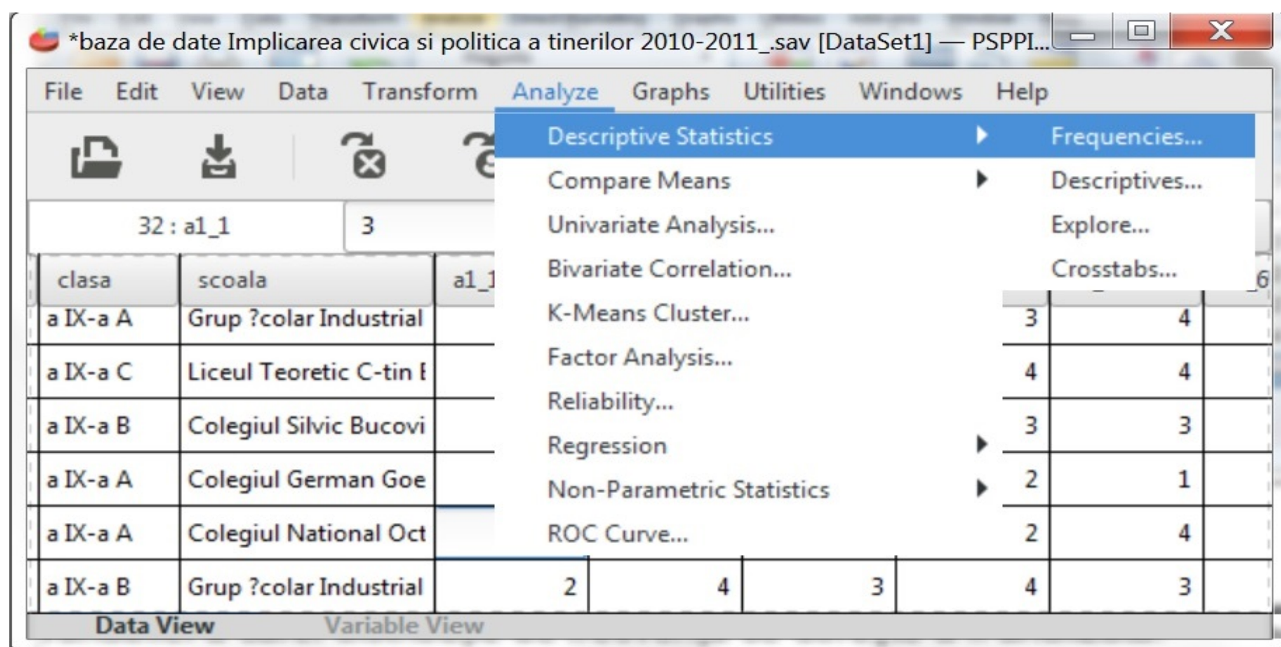
Graficul 1.14 Lansarea calculului unei distribuții de frecvență în SPSS



Comenzile **Analyze, Descriptive Statistics, Frequencies** conduc la lansarea unei ferestre de dialog care permite selectarea variabilei a cărei distribuție de frecvență dorim s-o analizăm în SPSS.

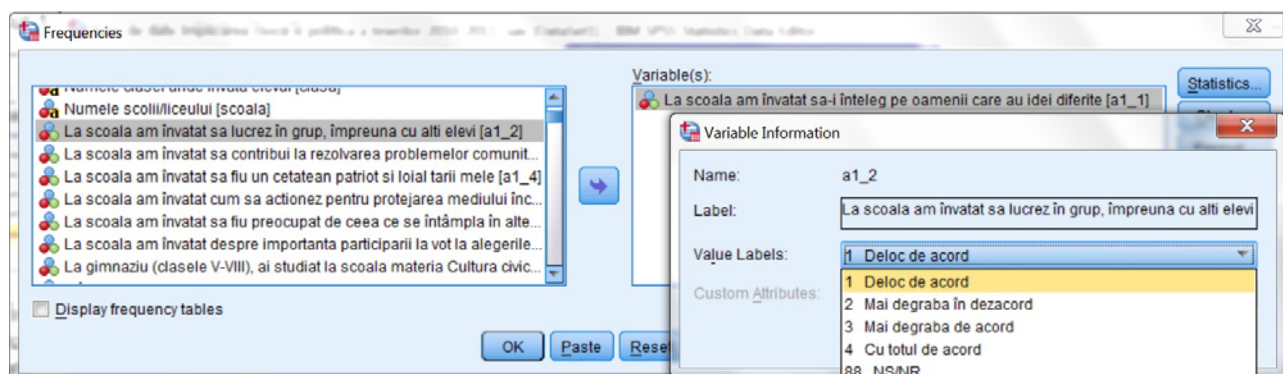
Pentru PSPP comenzile sunt identice dar procedurile statistice disponibile sunt mult mai limitate (Basic Stats in PSPP For NICAR 2016 Denver).

Graficul 1.15 Lansarea calculului unei distribuții de frecvență în PSPP



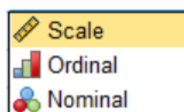
În fereastra de dialog din SPSS deschisă prin comanda Analyze, Descriptive Statistics, Frequencies, putem alege variabila pentru care dorim stabirea distribuției de frecvență. În plus, cu click dreapta deasupra uneia dintre variabile putem alege (similar cu programul PSPP) vizualizarea numelui sau etichetei veriabilelor precum și modul de ordonare al acestora.

Graficul 1.16 Selecția variabilei pentru calculul distribuției de frecvență SPSS



În graficul 1.16 e prezentată fereastra de dialog în care s-a optat pentru vizualizarea etichetelor variabilelor sortate în ordinea introducerii lor în baza de date. De remarcat, totuși, că la sfârșitul etichetei variabilei apare în paranteză și numele acesteia. Variabila a1_1, având eticheta „La scoala am învatat sa îi înțeleg pe oamenii care au idei diferite", a fost aleasă pentru a i se calcula distribuția de frecvență (click simplu deasupra variabilei în fereastra din stânga și apoi transferul respectivei variabile în căsuța din partea dreaptă, având titulatura „Variable(s):" Dacă dorim să aflăm informații despre alte variabile din baza de date, dăm click dreapta deasupra celei pe care am ales-o (a1_2 în graficul 1.16) și deschidem o nouă fereastră care oferă informații despre variabilă, cum ar fi: nume, eticheta acesteia și etichetele pentru valorile numerice ale caracteristicii din baza de date (Official IBM SPSS Statistics 21 Brief Guide).

De notat faptul că în partea dreaptă a căsuței de dialog pot fi selectate mai multe variabile și trimise în partea dreaptă pentru a li se calcula distribuția de frecvență utilizând butonul cu săgeată aflat la mijloc. În plus, e importamt de reținut că variabilele din SPSS au un simbol înaintea numelui care indică natura sociologică a acestora:



- Cantitativă
- Calitativă ordinală
- Calitativă nominală

Privitor la programul PSPP mai punem în evidență doar că și aici vizualizarea caracteristicilor include utilizarea unui simbol (a se vedea graficul 1.13)

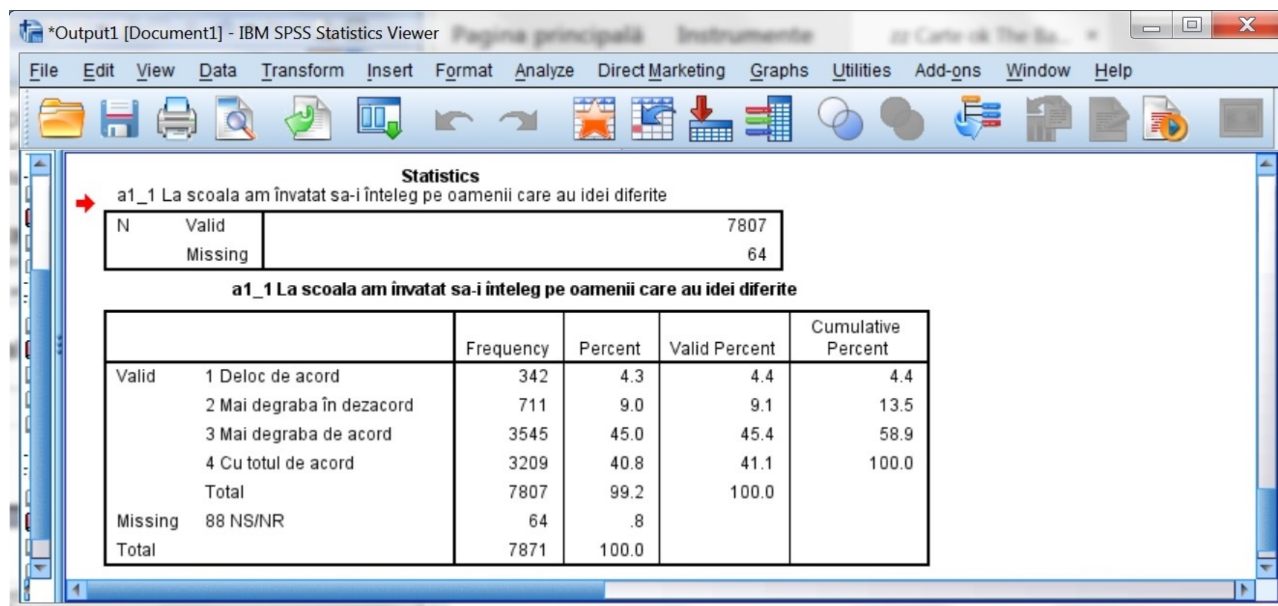


- Calitativă nominală
- Calitativă ordinală
- Cantitativă

În rest, fereastra de dialog PSPP pentru selectarea variabilei ce va fi inclusă în analiza frevenței nu oferă facilitățile existente în SPSS (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

Cu click pe butonul OK (graficul 1.16) obținem fereastra output (extensia .spv) care prezintă distribuția de frecvență a variabilei. Primul tabel indică totalul răspunsurilor valide din baza de date (7807) și datele lipsă (64).

Graficul 1.17 Tabelul de frecvență al variabilei în SPSS



Statistics

a1_1 La scoala am învatat sa-i înteleg pe oamenii care au idei diferite

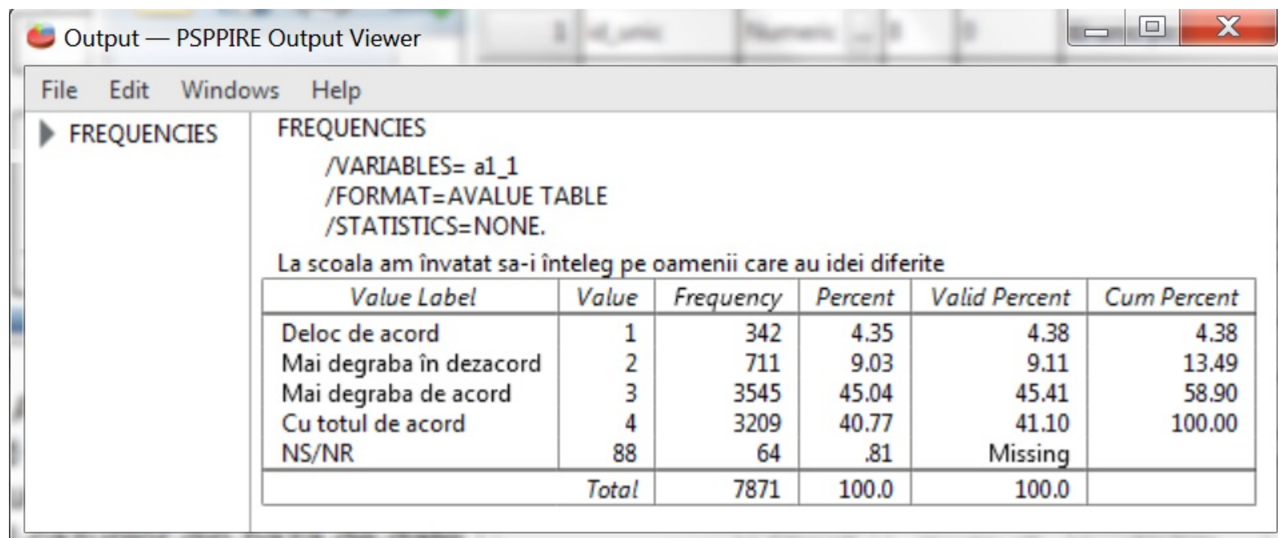
N	Valid	7807
	Missing	64

a1_1 La scoala am învatat sa-i înteleg pe oamenii care au idei diferite

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Deloc de acord	342	4.3	4.4	4.4
	2 Mai degraba în dezacord	711	9.0	9.1	13.5
	3 Mai degraba de acord	3545	45.0	45.4	58.9
	4 Cu totul de acord	3209	40.8	41.1	100.0
	Total	7807	99.2	100.0	
Missing	88 NS/NR	64	.8		
Total		7871	100.0		

Al doilea tabel prezintă efectiv distribuția de frecvență, în prima coloană având valorile variabilei, urmate de etichetele aceatora. În a doua coloană (Frequency) se prezintă frecvențele proprii diferitelor valori, prin raportare la totalul cazurilor din baza de date. În coloana Valid Pertcent sunt prezentatele procentele așa-ziselor răspunsuri valide (unde raportarea se face doar la totalul celor care și-au exprimat părerea / au oferit informația solicitată). În ultima coloană a tabelului se prezintă procentele cumulate, care indică ponderea celor care au formulat unul dintr-o serie de răspunsuri succesive posibile (de exemplu, 13.5% dintre respondenți au răspuns „Deloc de acord" sau „Mai degrabă în dezacord" (SPSS SURVIVAL MANUAL A step by step guide to data analysis).

Graficul 1.18 Tabelul de frecvență al variabilei în PSPP




```
FREQUENCIES
    /VARIABLES= a1_1
    /FORMAT=AVALUE TABLE
    /STATISTICS=NONE.
```

La scoala am învatat sa-i înteleg pe oamenii care au idei diferite

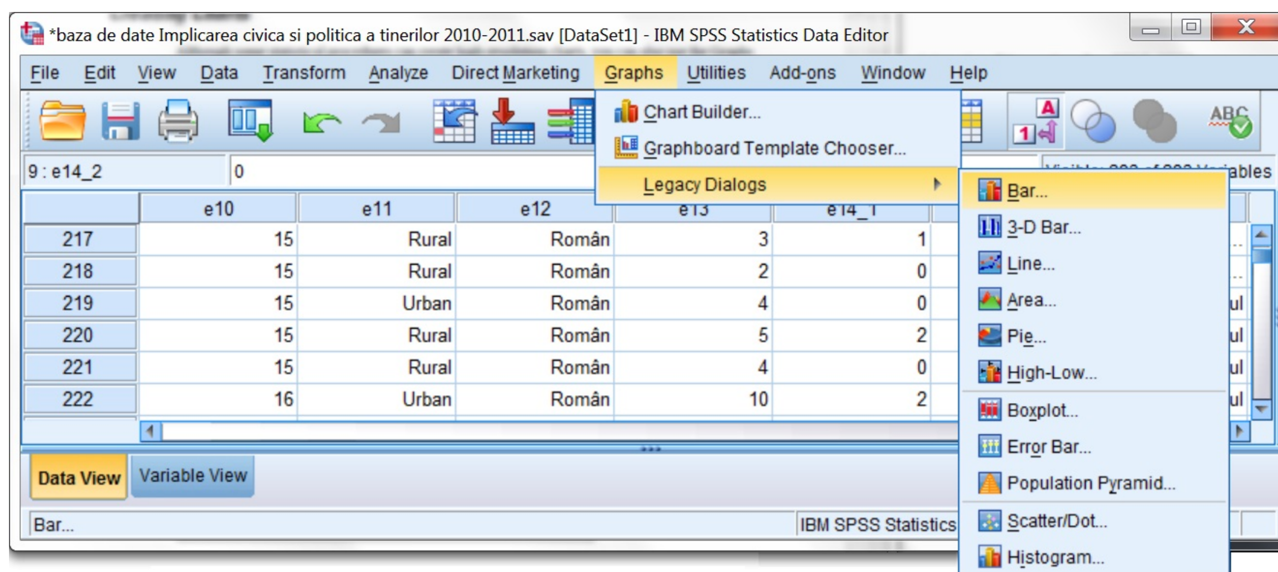
Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Deloc de acord	1	342	4.35	4.38	4.38
Mai degraba în dezacord	2	711	9.03	9.11	13.49
Mai degraba de acord	3	3545	45.04	45.41	58.90
Cu totul de acord	4	3209	40.77	41.10	100.00
NS/NR	88	64	.81	Missing	
	Total	7871	100.0	100.0	

În graficul 1.18 prezentarea datelor e apropiată de cea din programul SPSS, de această dată existând un singur tabel în care prima coloană reprezintă etichetele valorilor variabilei iar cea de a doua, valorile acesteia. Urmează apoi frecvențele, procentele obținute prin raportare la totalul cazurilor, procentele valide (obținute în raport cu datele valide) și, în final, procentele cumulate.

1.4 Crearea unor grafice în SPSS și PSPP

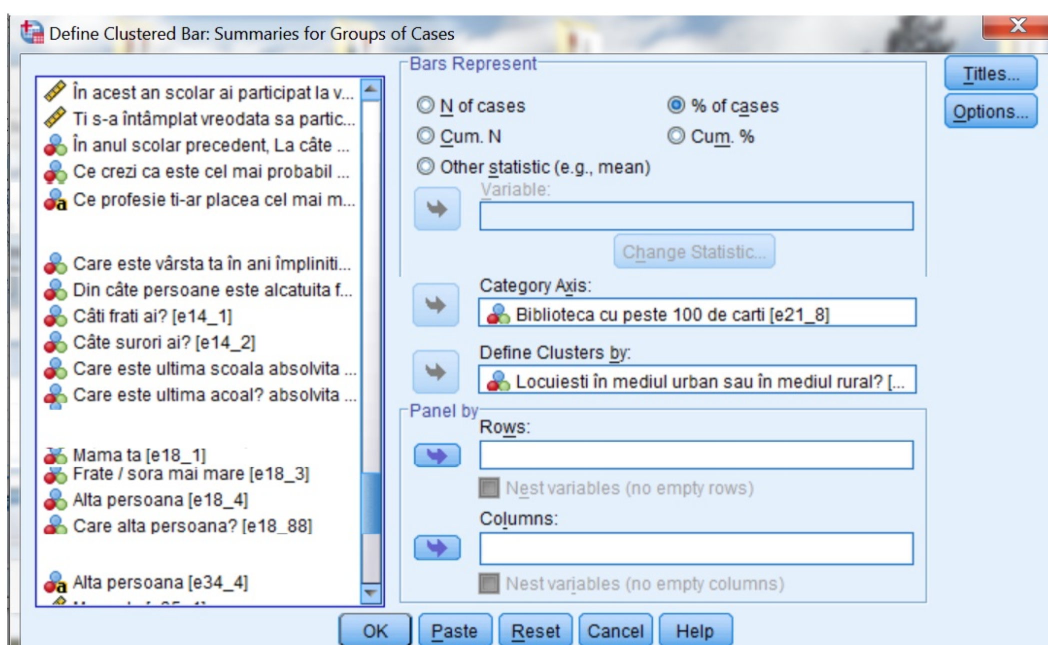
Pentru construirea unui grafic simplu putem opta pentru meniul **Graphs**, instructiunea **Legacy dialogs** și **Bar** (SPSS Manual for Introductory Applied Statistics).

Graficul 1.19 Utilizarea meniului Graphs pentru construirea graficelor SPSS



Vom obține o fereastră de dialog în care putem selecta variabilele de interes.

Graficul 1.20 fereastra de dialog pentru construcția graficelor SPSS



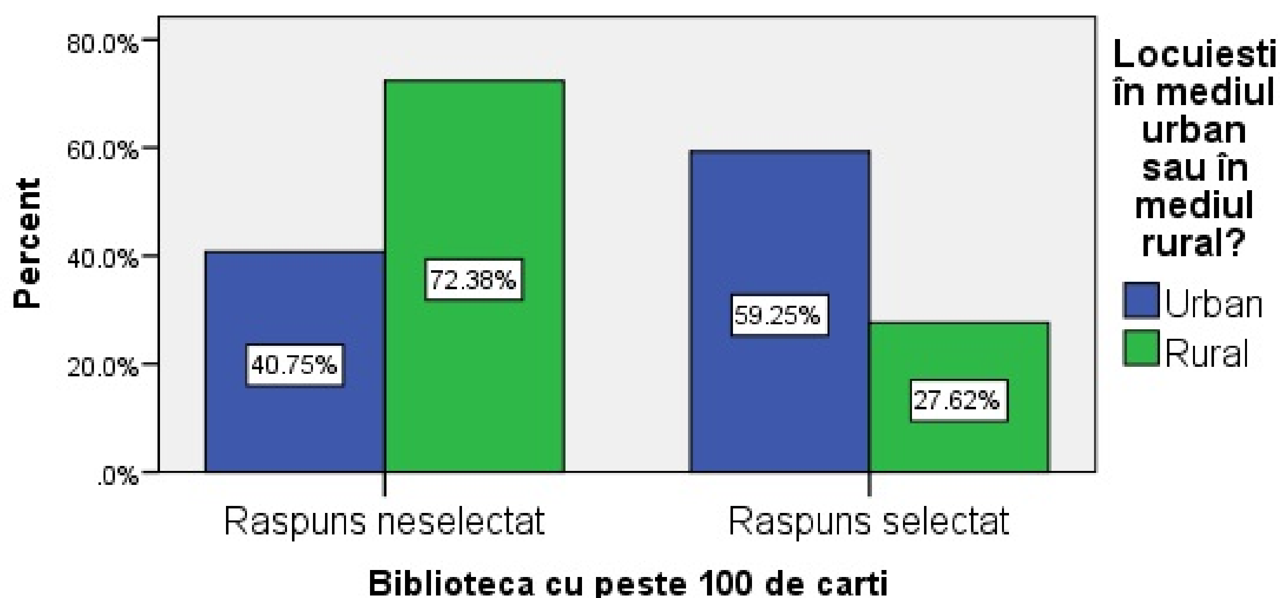
În partea de sus a ferestrei de dialog din graficul 1.20 se poate alege natura reprezentării grafice: număr de cazuri, procent, procent cumulat sau un alt tip de calcul (cum ar fi valorile medii).

În spațiul de selecție „Category Axis" putem indica variabila pe care dorim să o analizăm (în exemplul nostru, faptul de a poseda sau nu o bibliotecă cu cel puțin 100 de cărți).

În spațiul de selecție „Define clusters by" se va introduce caracteristica în ale cărei categorii se dorește analiza variabilei care a fost inițial selectată, adică posesia bibliotecii.

În partea de jos a ferestrei de dialog mai există două spații de selecție cu preambulul „Panel by" (Rows respectiv Colomns) ceea ce desemnează faptul că se pot utiliza variabile de control (pe lini sau coloane) care să permită analiza relației dintre cele două caracteristici definite inițial (folosind „Category Axis" și „Define clusters by") în mod separat, în categoriile respectivelor variabile de control (SPSS Manual for Introductory Applied Statistics).

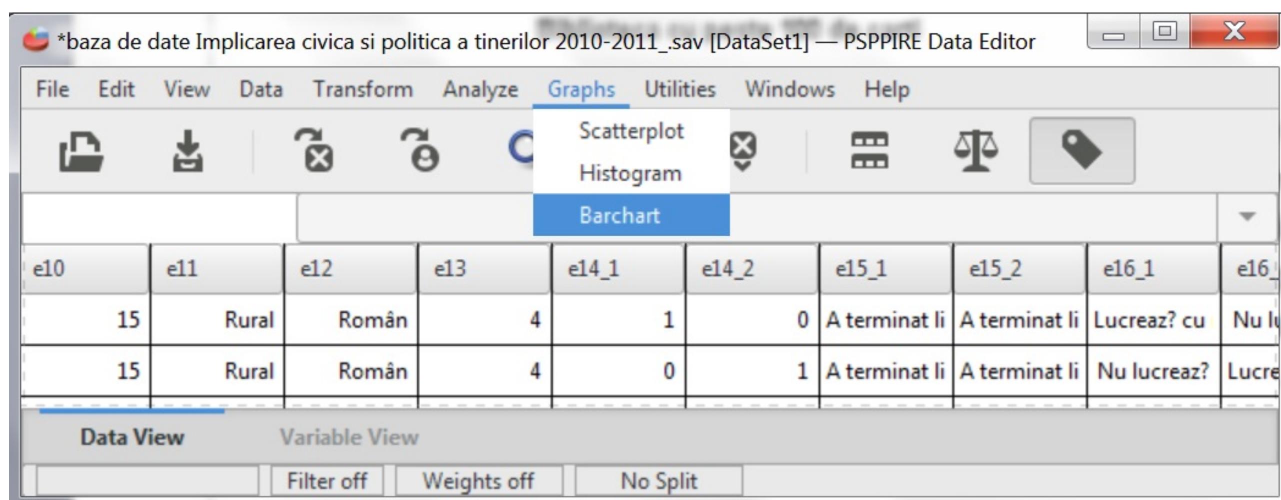
Graficul 1.21 Relația dintre rezidență și mărimea bibliotecii în SPSS



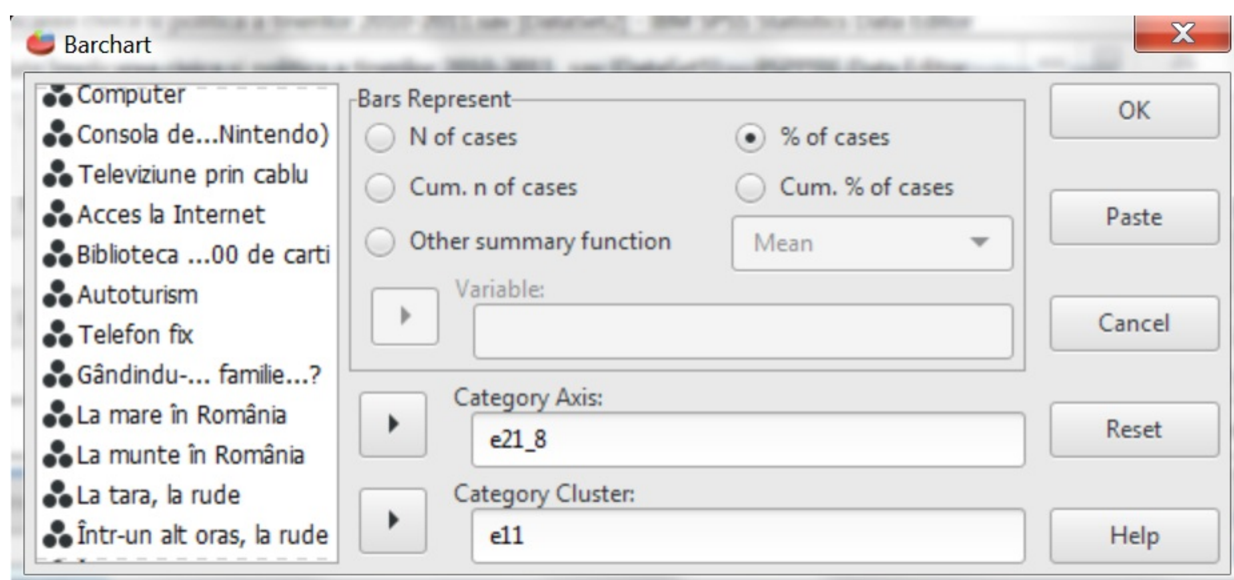
Pe axa orizontală (Category Axis: „Biblioteca cu peste 100 de cărți") a graficului 1.21 sunt grupați, la stânga, cei ce nu posedă cel puțin 100 de cărți respectiv cei ce posedă (dreapta). Fiecare sub-grup e divizat în funcție de rezidență, astfel încât putem observa că la nivelul mediului urban 59.25% din familii posedă cel puțin 100 de cărți pe când la sat doar 27.62% se află în această situație. Pe axa verticală a graficului apar procentele - opțiunea Bars represent (Using SPSS 20, Handout 3: Producing graphs).

Cu programul PSPP vom apela tot la meniul Graphs pentru a construi un grafic, în acest caz opțiunile fiind sensibil mai limitate decât în SPSS. Prin succesiunea de comenzi **Graphs**, **Barchart**, (a se vedea graficul 1.22) vom obține o fereastră de dialog asemănătoare cu cea din programul SPSS dar căreia îi lipsesc opțiuniele de introducere a variabilelor de control pe linii sau pe coloane (vezi graficul 1.23). În plus, procentele din Graficul 1.24 (care nu pot fi vizualizate) sunt calculate în raport cu totalul respondenților (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1). Graficul 1.24 prezintă dezavntajul imposibilității compării directe a celor două medii de rezidență (dacă, evident, sunt diferențe importante de volum al populației între rural și urban).

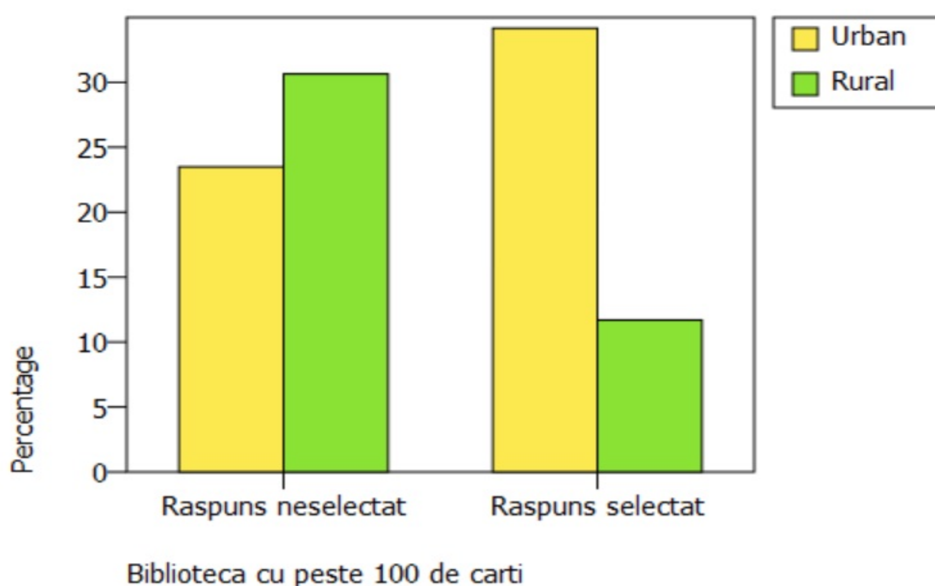
Graficul 1.22 Utilizarea meniului Graphs pentru construirea graficelor



Graficul 1.23 Fereastra de dialog pentru construcția graficelor PSPP



Graficul 1.24 Relația dintre rezidență și mărimea bibliotecii în PSPP



1.5 Părăsirea programului SPSS respectiv PSPP

În momentul în care dorim să părăsim programul SPSS trebuie să ne asigurăm că am salvat modificările pe care le-am făcut fisierului tip bază de date (cu extensia .sav) și că am indicat corect folderul în care să fie păstrate fișierele de tip output (cu extensia .spv).

De reținut faptul că programul SPSS nu face salvări automate, astfel încât orice informație va fi pierdută prin închiderea ferestrei Data Editor (care conține meniurile principale - File, Edit, Wiew, etc.) dacă, în prealabil, fișierele create sau modificate nu au fost salvate prin succesiunea File, Save as, indicând calea de urmat spre folderul de destinație și, dacă e cazul, tipul de fisier (A Handbook of Statistical Analyses using SPSS).

Observațiile de mai sus sunt valabile și pentru programul PSPP, cu precizarea că fisierele de tip output din SPSS pot fi gestionate si folosind selecția unuia sau mai multor obiecte (grafice, tabele) cu copy urmat de past, inclusiv într-un fisier Word, pe când analizele de date realizate în PSPP pot fi salvate doar global, sub forma unui fisier pdf sau html. Există, totuși, o excepție, tabelele (de frecvență, de asociere, etc.) pot fi selectate folosind succesiunea de comenzi **Edit, Copy** (după ce în prealabil s-a selectat numele tabelului prezent în partea stângă a ferestrei ce conține rezultatele analizei statistice). Se poate da comanda **Paste** într-un fisier word dar tabelele obținute nu vor fi de tip Microsoft Word ci succesiuni de litere, cifre și linii (orizontale sau verticale) care vor recompune tabelul inițial (dacă se va păstra tipul de font inițial, adică Lucida Console).

Tipurile de fisiere care pot fi alese pentru salvarea bazelor de date din SPSS și PSPP (identice între ele, cu extensia.sav) sunt aceleași cu fisierele ce pot fi deschise prin intermediul acestor programe. SPSS permite mult mai multe variante de schimbare a naturii (extensiei) fisierelor bază de date (.sav) decât PSPP.

După scurta trecere în revistă (în termeni comparativi) a facilităților oferite de programele SPSS și PSPP, ne vom ocupa de prezentarea mai sistematică a acestora, analizând, în capitolul următor, sursele și modul de organizare a datelor.

2. SURSELE ȘI ORGANIZAREA DATELOR ÎN SPSS ȘI PSPP

2.1 Surse de date pentru programele SPSS și PSPP

Vom începe cu o listă a posibilelor surse de date, așa cu apare în lucararea The Basics: SPSS for Windows (2002, p. 36):

- „interviuri
- chestionare
- observații
- evaluari (de către colegi, personal, experți)
- teste psihometrice de personalitate, valori, norme, atitudini, credințe
- scripte, arhive ale instituțiilor
- statisticile guvernamentale
- teste privind nivelul de informare, teste de interpretare, competențe, aplicare a cunoștințelor
- teste proiective
- teste situaționale care prezintă respondentului situații de viață simulate
- înregistrări din jurnalele unor persoane
- dovezi fizice
- examene clinice
- evidențe financiare
- diferite documente scrise (procese verbale ale ședințelor consiliului de administrație, ziare, transcrieri, etc.)”

Autorii arată că lista nu e nici pe departe exhaustivă, având ca scop doar să ofere o imagine asupra potențialelor surse de date. Obiectivele cercetării, resursele disponibile și o mare doză de ingeniozitate și imaginație vor fi călăuzele în selectarea metodologie cercetării.

2.2 Organizarea datelor în SPSS și PSPP

După cum am menționat deja, modul în care sunt organizate datele în SPSS poate fi vizualizat cu ajutorul ferestrei **Data Editor,** obținută prin lansarea programului SPSS și deschiderea unui fișier tip bază de date (având extensia .sav). În partea din stânga jos a respectivei ferestre putem selecta „Data view”, ceea ce semnifică prezentarea datelor existente în baza de date sau „Variable view”, opțiune ce oferă informații exhustive despre variabilele care au fost definite în respectiva baza de date (SPSS Manual for Introductory Applied Statistics: A Variable Approach).

În Graficul 2.1 am ales să vizualizăm etichetele valorilor efectiv introduse în baza de date (selectând opțiunea „Value labels” din meniul „Wiew”).

Fiecare linie din tabelul existent în fereastra „Data editor” corespunde unei unități statistice din baza de date. Ar putea fi vorba despre oameni, localități, firme sau orice entități ce formează o populație în sens statistic. Singura condiție este ca unitățile statistice să aibă o anumită caracteristică comună, prin care să le putem delimita drept țintă a studiului nostru și să se diferențieze din perspectiva valorilor luate în raport cu o serie de caracteristici (variabile) care vor face obiectul analizei.

Prima coloană din tabel reprezintă, pur și simplu, un număr de ordine alocat liniilor acestuia, care nu se află în relație cu o anumită unitate statistică (unitățile statistice din tabel pot fi ordonate după diferite criterii, schimbându-li-se, astfel, numărul de ordine). Coloanele care urmează reprezintă caracteristici (variabile) definite la nivelul populației statistice.

Citind pe rânduri tabelul (prezentat în graficul 2.1), putem spune că individul corespunzând actualmente celei de a 1470-a înregistrare din baza de date studiază în București, la școala cu codul 10667, în clasa a VIII-a C, având titulatura de „Școala generală nr. 97" și că este „cu totul de acord" cu enunțul „La școală am învățat să-i înțeleg pe oamenii care au idei diferite".

Reamintim faptul că eticheta unei variabile poate fi vizualizată poziționând mouse-ul deasupra butonului din parta de sus a coloanei care conține numele variabilei. În exemplul prezentat în graficul 2.1 a fost vizualizată eticheta variabilei cu numele a1_1, adică „La școală am învățat să-i înțeleg pe oamenii care au idei diferite". Răspunsurile pe aceeași linie pentru întrebările a1_2 și a1_3 au fost „cu totu de acord" respectiv „mai degrabă de acord" (întrebarea 1_2 a fost următoarea: „La şcoală am învăţat să lucrez în grup, împreună cu alţi elevi"; formularea întrebării a1_3 a fost: „La şcoală am învăţat să contribui la rezolvarea problemelor comunităţii"). A se vedea „SPSS Manual for Introductory Applied Statistics: A Variable Approach".

Graficul 2.1 Organizarea datelor în SPSS. Data editor, varianta Data view

*baza de date Implicarea civica si politica a tinerilor 2010-2011.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 293 of 293 Variables

La scoala am invatat sa-i înteleg pe oamenii care au idei diferite

	localitate	codScoala	clasa	scoala	a1_1	a1_2	a1_3
1468	Zalau	11864	a IX-a G	grup scolar voievodul ...	Cu totul de acord	Mai degraba de...	Cu totul de acord
1469	Tg. Frumos	12904	a VIII-a C	ion creanga	Mai degraba de...	Cu totul de acord	Mai degraba de...
1470	Bucuresti	10667	a VIII-a C	scoala generala nr. 97	Cu totul de acord	Cu totul de acord	Mai degraba de...
1471	Letcani	8322	a VIII-a B	scoala cu clasele i-viii...	Mai degraba în ...	Cu totul de acord	Mai degraba în ...
1472	Macin	12849	a XII-a A	liceul gheorghe m.mur...	Mai degraba de...	Cu totul de acord	NS/NR
1473	Salonta	13086	a XI-a F	colegiul national aran...	Cu totul de acord	Mai degraba de...	Mai degraba în ...

Data View Variable View

IBM SPSS Statistics Processor is ready

Citind tabelul pe coloane vedem, în primul rând, din ce localități provin elevii care actualmente au numerele de ordine din baza de date între cuprinse între 1468 și 1473 (Zalău, Tg. Frumos, București, Lețcani, Măcin și Salonta). Este vorba, de fapt de o variabilă (caracteristică) definită la nivelul eşantionului de elevi şi care face să corespundă fiecărui respondent localiatea în care își desfășoară studiile. Analog, variabila „codScoala" face să corespundă oricărui elev codul școlii în care studiază, variabila „clasa" pune în relație respondenții cu numele clasei de care aparțin iar „scoala" precizează numele instituției în care studiază. Ultimele variabile indică răspunsurile elevilor la trei întrebări vizând anumite aspecte învățate la școală (Using SPSS For Windows Data Analysis and Graph).

Analizând graficul 2.2, care ilustrează fereastra „Data Editor” (în varianta de prezentare a datelor) din programul PSPP, constatăm că nu există diferențe notabile față de graficul 2.1, discutat anterior. Lipsește prima coloană care realiza o numerotare a liniilor din baza de date dar acest fapt nu are, practic, nici o relevanță pentru modul de organizare a datelor. Nici o procedură SPSS nu utilizează numerotarea din prima coloană „Data editor” pentru prelucrări statistice. (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

Graficul 2.2 Organizarea datelor în PSPP. Data editor, varianta Data view

*baza de date Implicarea civica si politica a tinerilor 2010-2011_.sav [DataSet1] — PSPPIRE Data Editor

File Edit View Data Transform Analyze Graphs Utilities Windows Help

1473 : judet | Bihor

	localitate	codScoala	clasa	scoala	a1_1	a1_2	a1_3
	Zalau	11864	a IX-a G				totul de acord
	Tg. Frumos	12904	a VIII-a C	ion creanga	Mai degraba de ac	Cu totul de acord	Mai degraba de ac
	Bucuresti	10667	a VIII-a C	scoala generala nr. 97	Cu totul de acord	Cu totul de acord	Mai degraba de ac
	Letcani	8322	a VIII-a B	scoala cu clasele i-viii	Mai degraba în de:	Cu totul de acord	Mai degraba în de:
	Macin	12849	a XII-a A	liceul gheorghe m.mu	Mai degraba de ac	Cu totul de acord	NS/NR
	Salonta	13086	a XI-a F	colegiul national aran	Cu totul de acord	Mai degraba de a	Mai degraba în de:

La scoala am învatat sa-i înteleg pe oamenii care au idei diferite

Data View | Variable View

Filter off | Weights off | No Split

Este important să nu se facă o confuzie între întrebările formulate într-un chestionar și caracteristicile definite în baza de date din SPSS sau PSPP. Adeseori, pentru aceeași întrebare trebuie construite mai multe variabile, cum ar fi cazul, care poate fi considerat clasic, al întrebărilor cu răspunsurile multiple. Dacă elevii ar fi fost întrebați în ce clasă au practicat un sport de performanță (a VIII-a, a IX, a X-a, a XI-a și/sau a XII-a), pentru fiecare dintre cei cinci ani de studiu trebuia creată o variabilă în baza de date care să ia valoarea 1 dacă în respectivul an respondentul a practicat sport de performanță și 0 în caz contrar. Valorile de 1 și 0 au fost alese în mod arbitrar, dar conferă caracteristicilor dihotomice proprietățile variabilelor dummy, care, deși calitative, pot fi tratate statistic identic cu cele cantitative. Argumentul ar fi că valoarea medie are sens să fie calculată deoarece reprezintă chiar ponderea celor ce au declarat, în cazul nostru, că au practicat sport de performanță în clasa de referință. Pe de altă parte, există întebări generale, care vizează mai multe aspecte:

„A1. Am dori să aflăm câteva lucruri despre ceea ce ai învăţat la şcoală. Pentru fiecare dintre afirmaţiile de mai jos, eşti sau nu de acord că se potrivesc în cazul tău?”

Respondentul va trebui, mai departe, să dea un răspuns pentru fiecare afirmație, arătându-și nivelul de acord / dezacord și pentru fiecare răspuns va trebui creată o variabilă specifică în baza de date (ex.: a1_1, a1_2 și a1_3)

Uneori apare un număr mai mare de variabile în baza de date decât cele corespunzând chestionarului aplicat inițial deoarece o anumită întrebare este aplicată aceleiași populații (aceluiați eșantion) la momente diferite de timp. Astfel de studii se realizează, de exemplu, atunci când se dorește evaluarea schimbărilor de atitudinale sau de comportament care a au fost promovate printr-o politică socială (îmbunătățirea obiceiurilor alimentare, diminuarea consumului de alcool, tutun, droguri, o conduită de prezervare a mediului natural, etc.). Comparația dintre răspunsul dat la primul și al doilea moment de timp oferă informații pentru schimbările care au avut loc la nivelul fiecărui individ în respectiva perioadă dar, mai ales, permite evaluarea globală a eficienței cu care s-a derulat campania de promovare a sănătății. A se vedea, pentru mai multe detalii, „Analysis of Multiple-Response Data".

2.3 Definirea unei baze de date în SPSS și PSPP

Adeseori ne aflăm în situația de a dispune de date statistice provenite dintr-o anchetă sociologică, din statisticile unor firme pe care le-am avut ca țintă în cercetare sau de la instituții specializate în colectarea informațiilor („Institutul Național de Statistică", „EUROSTAT", „The European Statistical System", etc.). Problema care se pune este de a aduce aceste date sub o formă cât mai convenabilă pentru a putea fi analizate statistic cu soft-urile specializate care ne interesează (SPSS și PSPP). Există o multitudine de modalități prin care putem introduce date statistice în SPSS datorită posibilității acestui program de a accesa baze de date care au fost create în soft-uri cum ar fi Excel (extensia .xls, .xlsx sau .xlsm), Lotus (extensia .w), Sylk (exensia .slk), dBase (extensia .dbf), etc., după cum se poate observa în graficul 1.7 (a se vedea „SPSS for Windows Version 19.0: A Basic Tutorial").

Situația este sensibil diferită pentru PSPP, unde nu se pot prelua decât baze date având extensia .sav (create în același program sau în SPSS) respectiv fisiere Portable, cu extensia .por, utilizate în mod specific pentru transferul datelor (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

În această secțiune ne vom ocupa de preluarea datelor în programele SPSS și PSPP printr-o modalitate clasică, adică prin introducerea lor de la tastatură, în condițiile în care nu dispunem de un format electronic al informației care să poată fi interpretat de soft-urile dedicate analizei statistice (în mod tipic, chestonarele care au fost aplicate unui eșantion de respondenți dar și informațiile despre agenții economici existente online sau datele vizând localitățile României cuprinse în anuarele statistice). Pentru răspunsurile la fiecare întrebare (și, în general, pentru fiecare informație pe care dorim s-o transferăm în SPSS sau PSPP) va trebui să definim modul potrivit de a realiza codificarea. Aceasta se realizează, în general, sub formă numerică, pentru simplitate, în sensul că unei variante de răspuns îi va corespunde una sau mai multe cifre (SPSS SURVIVAL MANUAL A step by step guide to data analysis using SPSS for Windows).

Până în acest moment, am utilizat varianta de prezentare a programelor SPSS și PSPP sub forma „Data Editor, Data View", care permite vizualizarea valorilor introduse în baza de date sau a etichetelor acestor valori (a se vedea graficele 1.9, 1.10 și 1.11). În continuare dorim să arătăm care este modalitatea de a construi o bază de date potrivită pentru introducerea informațiilor de natură statistică pe care le deținem. În acest scop vom utiliza tot fereastra (domeniul) „Data Editor" dar în varianta „Variable View", care e destinată definirii variabilelor existente în baza de date (tipul variabilei, număr maxim de caractere, etichete ale variabilelor și valorilor acestora, etc.).

Să presupunem că dorim să introducem în SPSS (PSPP) informațiile culese cu ajutorul unor chestionare aplicate elevilor în care am fost interesați de reușita lor școlară, de timpul alocat pregătirii pentru școală, de importanța acordată școlii de către tineri, de studiile părinților acestora și de modul în care ar caracteriza, printr-un singur cuvânt, propria școală.

Secvență de chestionar adresată elevilor (rol de exemplificare)

Numărul chestionarului ___

1. Cu ce medie generală ai terminat anul trecut școlar? ____

2. Cam cât tim aloci studiului într-o zi obișnuită de școală?
1) sub 15 min. 2) 15-29 min. 3) 30-59 min. 4) 1-1.99 ore 5) 2-2.99 ore 6) 3 ore sau peste

3. Cât de importantă ți se pare școala pentru reușita în viață?
1) Deloc 2) Puțin 3) Așa și așa 4) Importantă 5) Foarte importantă 8) Nu știu

4. Care e nivelul de studii absolvit de părinții tăi?
Mama: 1) max. 8 clase 2) profesională 3) liceu 4) postliceală 5) studii superioare
Tata: 1) max. 8 clase 2) profesională 3) liceu 4) postliceală 5) studii superioare

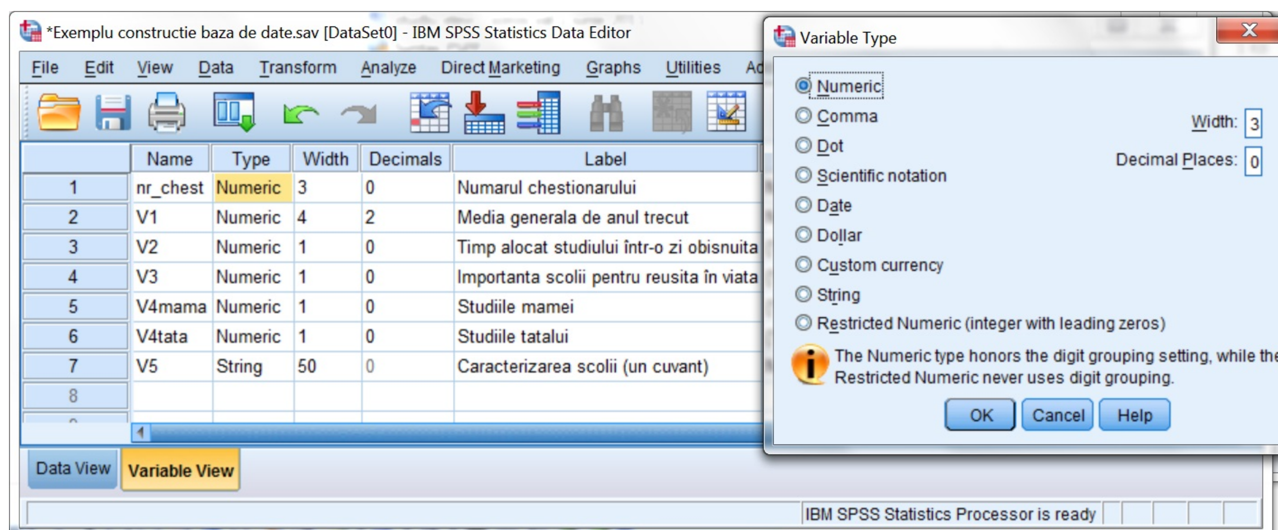
5. Cum ai caracteriza, într-un singul cuvânt, școala ta?
____

Pentru a putea introduce informațiile în baza de date va trebui, în primul rând, să stabilim numele fiecărei variabile, apoi simbolul pe care îl vom folosi pentru variantele pre-definite de răspuns ale fiecărei întrebări (evident, dacă există). Va trebui, apoi, să alegem modul în care vom semnala datele lipsă, Frecvent sunt folosite numerele 9 sau 99 pentru a desemna lipsa informațiilor din chestionar sau, în general, din sursa de date pe care o folosim (a se consulta „SPSS Step-by-Step Tutorial - Part 1”).

Inițierea definirii bazei de date o putem face lansând programul în mod clasic: „Start”, „Program Files”, urmat de numele programului. Pentru SPSS va fi lansat domeniul „Data Editor” și o fererastră de dialog (a se vedea graficul 1.1) în care va trebui să selectăm, din partea dreaptă, opțiunea „Type in data”. (același rezultat îl obținem, însă, și dacă închidem, pur și simplu, respectiva fereastră de dialog). Vom alege, în continuare, varianta „Variable view” din partea de jos a ferestrei „Data Editor” pentru a putea să definim baza de date necesară introducerii datelor culese cu secvența de chestionar prezentată mai sus.

Vom începe prin a tasta numele primei variabile, de exemplu prin literele „nr_chest” (de notat că în SPSS și în PSPP se pt utiliza nume relativ lungi pentru variabile – aproximativ 50 de caractere – dar acestea trebuie să înceapă cu o literă, nu trebuie să conțină spații sau simboluri speciale, cum ar fi virgula, semnul plus (+), egal (=) slash (/), semnul întrebării, punct și virgulă, două puncte, etc.). În interiorul numelui variabilei se poate folosi liniuța din partea de jos (_), recomnadată pentru a delimita cuvinte, sau chiar simbolurile @ respectiv $ (Introduction to SPSS - version 23 - for Windows, Practical workbook).

Graficul 2.3 Definirea unei baze de datee în SPSS



În graficul 2.3 se poate observa concordanța dintre între întrebările din chestionar formulate anterior și variabilele care au fost definite în baza de date a programului SPSS (evident, păstrarea ordinii întrebărilor este o opțiune posibilă, utilă pentru sistematizarea introducerii datelor, fără a avea, însă, un caracter obligatoriu).

În colana „**Name**” apar numele alese pentru variabile, care pot sugera conținutul respectivei întrebări („nr_chest” se referă, evident, la numărul atribuit chestionarului) sau pot fi simple simboluri care să arate clar la care întrebare din chestionar se referă (de exemplu, „V1” vizează întrebarea: „Cu ce medie generală ai terminat anul trecut școlar?”, care, în chestionarul aplicat tinerilor, era precedată de simbolul „V1”). În cazul întrebării „Care e nivelul de studii absolvit de părinții tăi?” au trebuit construite două variabile în baza de date, câte una pentru fiecare dintre părinți. Numele variabilei se scrie de la tastatură după ce, în prealabil, a fost selectată una dintre căsuțele coloanei „Name” (IBM SPSS Statistics 23, Brief Guide).

În coloana „**Type**”, dând click în căsuța din dreptul variabilei pe care o vizăm, vom obține o fereastră de dalog care permite selectarea tipului de date care vor fi introduse în baza SPSS, a numărului maxim de caractere care pot fi introduse pentru valorile variabilei precum și a numărului maxim de zecimale admis (dacă numărul e întreg se va trece 0 pentru zecimale).

Fereastra de dialog mai sus amintită (vizibilă în partea dreaptă a graficului 2.3) conține o listă cu posibile modalități de introducere a datelor în programul SPSS. Nu le vom discuta pe toate ci doar pe cele care sunt mai frecvent folosite (SPSS 12.0 Brief Guide).

Numeric desemnează introducerea datelor sub formă de numere, în modalitatea standard, cu sau fără zecimale (numărul maxim acceptat de caractere, – având titulatura „Width” – respectiv numărul maxim de zecimale – simbolizat prin „Decimal places” – se stabilesc din căsuțele din dreapta ferestrei de dialog). În graficul 2.3, pentru variabila având numele „nr_chest” (și eticheta „Numărul chestionarului”), s-a ales să fie reprezentată prin cel mult trei caractere și fără nici o zecimală, ceea ce înseamnă că valorile care pot fi introduse aparțin mulțimii numerelor întregi cuprinse între 0 și 999. Pentru media generală din anul trecut (având numele „V1”) se pot introduce numere cu partea întreagă având cel mult 2 cifre și cu patea zecimală respectând aceeași cerință (astfel, cel mai mare număr va fi 99.99). A se vedea și lucrarea „Using SPSS to Understand Research and Data Analysis”.

Tipul de date **„Comma”** (Opariuc-Dan, „Statistică aplicată în științele socioumane” 2009) vizează tot numere ce pot fi introduse similar variantei „Numeric”, singura diferență fiind folosirea, la vizualizarea datelor, a virgulei ca separator după cifra desemnând numărul miilor, al milioanelor, al miliardelor, etc. (punctul va rămâne separator pentru zecimale; de exemplu, numărul întreg 1000 va fi reprezentat 1,000.00, dacă pentru zecimale am selectat valoarea 2)

Tipul de date **„Dot”** pesupune același mod de introducere a datelor, diferența va consta în faptul că, la vizualizare punctul și virgula își schimbă rolurile (același număr întreg 1000 va fi vizualizat sub forma 1.000,00)

„Scientific notation” vizează prezentarea datelor (introduse, normal, sub formă numerică) folosind o notație științifică, care permite vizualizarea primele trei cifre ale numărului din baza de date sub forma unui întreg cu două zecimale, urmat de litera E ca separator pentru puterea lui 10 cu care tebuie multiplicat respectivul număr zecimal (obținând, astfel, o aproximare a valorii reale – care se păstrează, evident, nemodificată în baza de date). Păstrând același exemplu, numărul 1000 va fi reprezentat sub forma 1.00E3.

Varianta **„Date”** presupune alegerea unui format specific în care urmează a fi introdusă data calendristică și, implicit, un mod specific de vizualizare a informațiilor în baza de date. Modalitatea concretă de introducere / vizualizare a datelor va fi aleasă dintr-o fereastră de dialog care se va deschide în momentul în care se va da click pe varianta „date”.

„Dollar” vizează valori ale monedei americane pentru care se poate specifica formatul de prezentare și numărul maxim de caractere acceeptat folosind fereastra de dialog care apare când selectăm opțiunea „Dollar”. Valorile variabilei, vizualizate în baza de date, sunt precedate de simbolul $.

„Custom currency” ne permite să definim alte monede decât cele folosite în SUA dar datele se introduc sub forma clasică (și sunt prezentate identic cu varianta „Comma”, fiind vorba mai mult de o etichetare a datelor).

Alegând varinta „**String**” vom introduce și vizualiza datele sub fromă de șir de caractere (text). În general optăm pentru această variantă atunci când nu puem utiliza variante predeterminate de răspuns sau când formularea acestora ar fi ineficientă. De exemplu, pentru un eșantion național constuit simplu aleator, în care s-ar putea afla respondenți din oricare localitate a țării, ar fi absurd sa oferim variante predeterminate de răspuns. E mai firesc să folosim întrebări deschise, al căror răspuns va fi introdus ca atare în baza de date. În general, răspunsurile la orice întrebare deschisă sunt introduse, inițial, în format „String”, după care, dacă e cazul, sunt recodificate. SPSS-ul acceptă, pentru variabilele „string”, analizele statistice proprii caracteristicilor calitative. Pentru apartenența de sex putem folosi, fără probleme, simbolurile „M” și „F” în locul unora numerice. Totuși, precizăm că ori de câte ori e posibil, se preferă introducerea sub formă numerică a variantelor de răspuns alese de către respondenți pentru creșterea eficienței și scăderea riscului erorilor. Pentru variabila desemnând importanța școlii pentru reușita în viață e mai simlu să tastezi un număr decât să scrii varianta de răspuns (SPSS for Beginners, 1999, Vijay Gupta).

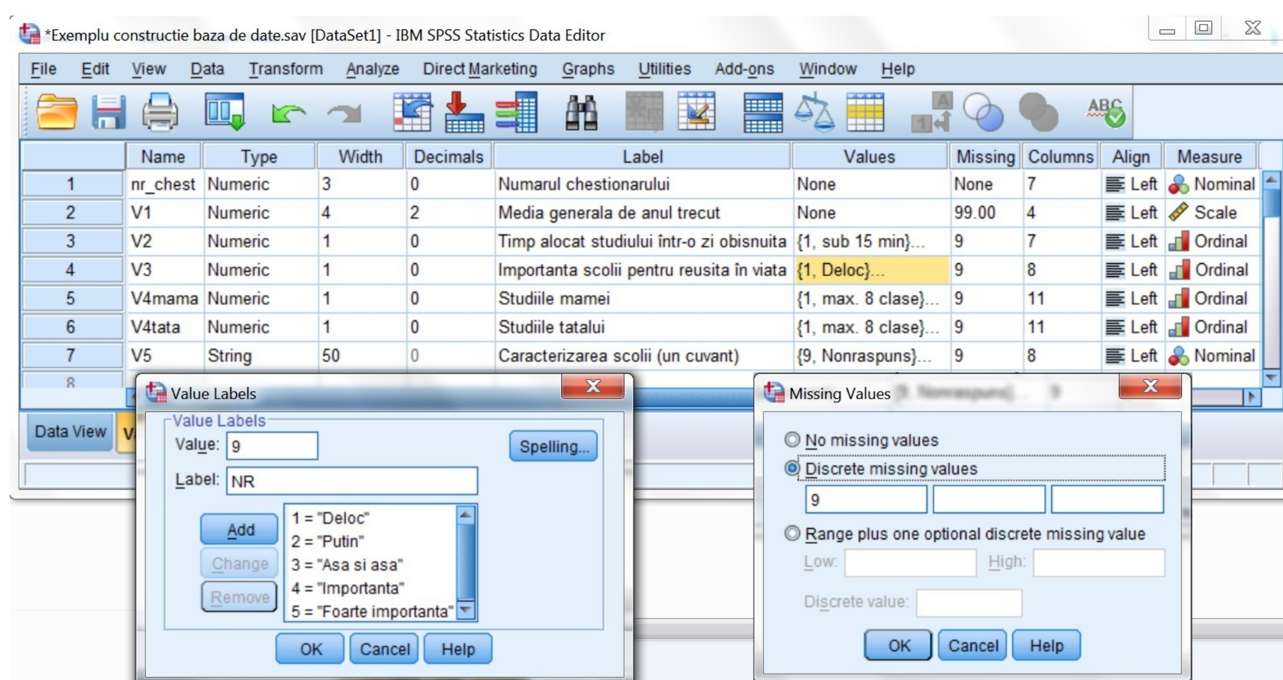
Revenind la fereastra principală „**Data Editor**”, în coloana **„Width”** se vor indica numarul maxim de caractere acceptate pentru introducerea valorilor unei caracteristici în baza de date din programul SPSS.

În coloana „**Decimals**” se vor alege, pentru variabilele numerice, numărul maxim de zecimale acceptate (în mod firesc, vor fi excluse variabilele de tip „Data” și „String”, unde zecimalele nu au sens).

Pentru orice tip de caracteristică putem defini, de asemenea, eticheta variabilei folosind opțiunea „**Label**”, care va permite specificarea semnificației numelui variabilei și va putea apărea în prelucrările statistice în locul sau alături de numele respectivei variabile.

În graficul 2.3, pentru variabila cu numele „V2” s-a stabilit eticheta „timp alocat studiului într-o zi obișnuită”, s-a optat pentru permiterea introcucerii unui singur caracter în baza de date (fără nici o zecimală, evident). În cazul variabilei „String” cu numele „V5” a fost atribuită eticheta „Caracterizarea școlii (un cuvânt)” și s-a limitat la maxim 50 de caractere lungimea unui șir de caractere (Analyzing data using SPSS, A practical guide).

Graficul 2.4 Definirea etichetelor valorilor și specificarea datelor lipsă în SPSS



În graficul 2.4 sunt prezentate ferestrele de dialog utilizate pentru definirea etichetelor variabilelor („**Value Labels**”) precum și a codurilor folosite pentru datele lipsă – „**Missing Values**” (nonrăspunsurile respondenților sau, în general, valorile care, din anumite motive, nu vor fi analizate împreună cu datele care sunt considerate valide).

Variabila cu numele „V3” este definită ca numerică în baza de date, este reprezentată printr-un singur caracter, fără zecimale, având eticheta numelui: „Timp alocat studiului într-o zi obișnuită”. Pentru o astfel de variabilă e foarte important să definim etichetele valorilor variabilei pentru a cunoaște semnificația numerelor introduse în baza de date. Dând click asupra căsuței care corespunde coloanei „Values” obținem ferastra de dialog „Value labels” (Using SPSS to Understand Research and Data Analysis).

Urmărind exemplul de chestionar prezentat anterior, în căsuța „Value” din fereastra de dialog „Value labels” vom introduce valoarea 1 iar în căsuța „Label” vom tasta eticheta acestei valori, adică „Puțin” (introducerea etichetei se va face, evident, de la tastatură). Urmează comanda „Add” pentru ca eticheta să fie realmente atribuită.

În graficul 2.4 este prezentat momentul în care se introduce ultima etichetă cu titulatura „NR” pentru valoarea 9 (cifra 9 semnifică faptul că nu s-a formulat un răspuns valid sau, altfel spus, simbolizează nonrăspunsurile subiecților pentru o anumită întrebare). După introducerea tuturor etichetelor valorilor este necesară comanda „OK” din partea de jos a ferestrei de dialog (Discovering Statistics using Spss).

Următorul pas în definirea variabilei „Timp alocat studiului într-o zi obișnuită” îl reprezintă definirea datelor lipsă, cu ajutorul ferestrei de dialog „**Missing Values**”, deschisă prin click în căsuța corespunzătoare coloanei „Missing” (atenție: cele două ferestre de dialog graficul 2.4 nu apar simultan ci succesiv; în figură e folosit un artificiu grafic pentru a simplifica prezentarea). A se vedea și ”Discovering Statistics using SPSS”

În coloana cu titulatura „**Columns**” din „Data Editor” (varianta „Variable View”) putem indica numărul de caractere care va fi fi vizualizat, pentru o anumită anumită variabilă, în momentul în care vom inspecta gestionarea datelor cu „Data Editor” însă în cealaltă varianta de prezentare, adică „Data View”.

De exemplu, pentru varabila cu numele V1, având eticheta „media generală din anul trecut”, sunt prevăzute patru caractere pentru vizualizarea variabilei (atunci când în fereastra „Data editor” sunt prezentate informațiile existente bazele de date, adică atunci când este activă varianta de prezentare a datelor cu titulatura „data view”), ceea ce este suficient având în vedere că nici numele variabilei, nici valorile acesteia nu depășesc numărul de caractere selectat pentru „Columns” (Analyzing data using SPSS).

Este important de notat că numărul pe care îl indicăm pentru fiecăre variabilă utilizând opțiuna „columns” nu influențează în nici un fel modul de definire sau reprezentare a variabilelor în baza de date SPSS ci stabilește, doar, numărul de caractere care vor fi vizibile.

Următoarea coloană, „**Align**”, se referă tot la vizualizarea datelor în ferestra „Data Editor”, varianta „Data View”, stabilind modul în care să fie aliniate datele în interiorul coloanei: la stânga, centrat sau la dreapta (Using SPSS to Understand Research and Data Analysis).

În coloana „**Measure**” se indică tipul variabilei din punct de vedere sociologic. „Nr_chest”, deși a fost definită ca variabilă numerică, ea nu indică nici intensitatea de manifestare a unui fenomen, nici nu ne permite să stabilim o relație de ordine între respondenți (bazată pe un criteriu cu relevanță sociologică). În consecință, va fi marcată ca fiind de tip nominal.

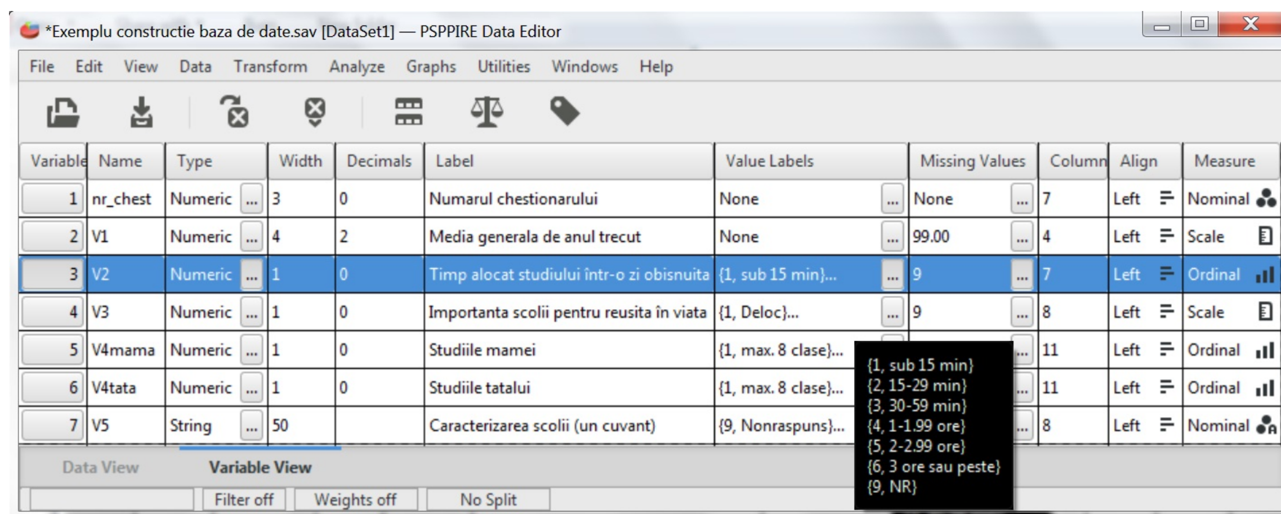
Pe de altă parte, variabila V2, cu eticheta „media generală de anu trecut” este o variabilă cantitativă și va fi încadrată ca „Scale”. Variabila „v3”, cu eticheta „importanța școlii pentru reușita în viață” este de tip ordinal (având variantele de răspuns „Deloc”, „Puțin”, „Așa și așa”, „Importantă”, „Foarte importantă”) și, ca atare, în coloana „measure” se va alege varianta „Ordinal”. A se consulta „Introduction to SPSS (version 16) for Windows”.

Încadrarea corectă a variabilelor din punct de vedere sociologic, în coloana „Measure” a ferestrei Data Editor (în varianta de prezentare „Variable View”), este importantă pentru eliminarea posibilelor confuzii privind natura cracteristicilor definite în baza de date (având în vedera faptul că, ori de câte ori este posibil, datele sunt introduse sub formă numerică, inclusiv pentru variabilele calitative, ordinale sau nominale).

Atragem atenția asupra faptului că o variabilă definită ca „String” (șir de caractere) în baza de date nu va putea fi declarată, în coloana „Measure”, ca fiind de tip „Scale” (în mod firesc, un șir de caractere nu va putea fi declarat ca având natura cantitativă). În rest, pentru variabilele definite ca fiind numerice, programul SPSS nu va impune limitări în definirea tipului sociologic al acestora.

Definirea unei baze de date în **PSPP** este extrem de asemănătoare cu cea din SPSS. Denumirile și semnificația coloanelor care permit indicarea naturii și proprietăților variabilelor sunt aceleași cu cele din SPSS, diferențele sunt la nivel de amănunte, legate de tastarea datelor sau vizualizarea lor (Basic Stats in PSPP For NICAR 2016).

Graficul 2.5 Definirea unei baze de date în PSPP



Alegerea numărului maxim de caractere pentru o anumită variabilă sau stabilirea zecimalelor acceptate se poate face prin tastare directă dar și utilizând butoanele „+” respectiv „-” care, prin click deasupra lor, determină creșterea sau descreșterea cu o unitate a respectivelor numere de caractere.

Dacă poziționăm mouse-ul deasupra unei căsuțe din coloana „Value Labels” va apărea un tabel care ne indica eticheta corespunzătoare pentru fiecare dintre valorile luate de variabila avută în vedere. Relevăm faptul că în graficul 2.5 se prezintă etichetele valorilor variabilei V2, desemnând „timpul alocat studiului într-o zi obișnuită”).

Dacă analizăm mai atent coloana „Measure” din „Data Editor” PSPP observăm că simbolul pentru caracteristicile nominale (din punct de vedere sociologic) diferă între variabila definită „Numeric” (nr_chest) și cea definită „String” (diferența constă în prezența literei „A” în simbolul utilizat pentru variabilele alfanumerice.

O ultimă diferență puțin semnificativă: în PSPP se semnalizează explicit faptul că variabilele care au fost definite ca „String” (în coloana „Type”) nu vor putea fi definite ca „Scale”, adică cantitative (apare imaginea „Scale Ø”). A se vedea „PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1”.

În continuarea vom prezenta, pe scurt, felul în care sunt introduse datele în programele SPSS și PSPP. Continuând exemplul formulat anterior, ne vom propune să introducem datele culese prin intermediul fragmentului de chestionar prezentat anterior pentru 7 elevi care au acceptat să răspundă la întrebări.

2.4 Introducerea valorilor variabilelor în bazele de date SPSS şi PSPP

Valorile variabilelor, pentru populația sau eșantionul cuprins în analiză, pot fi introduse în două modalități care sunt similare ca principiu dar care diferă ca tehnică sau eficiență (SPSS Survuval Manual, A step by step guide to data analysis using SPSS for Windows).

Graficul 2.6 Introducerea datelor în SPSS (Data Editor, Data View)

	nr_chest	V1	V2	V3	V4mama	V4tata	V5
1	1	9.42	4	4	4	5	Placuta
2	2	6.54	1	1	1	2	Plicticoasa
3	3	8.11	2	3	3	2	Banala
4	4	8.47	3	2	4	3	Obositoare
5	5	9.83	5	4	5	4	Interesanta
6	6	9.36	6	5	5	5	Incitanta
7	7	6.31	9	9	9	9	Inutila

	nr_chest	V1	V2	V3	V4mama	V4tata	V5
1	1	9.42	1-1.99 ore	Importanta	postliceala	studii supe...	Placuta
2	2	6.54	sub 15 min	Deloc	max. 8 clase	profesionala	Plicticoasa
3	3	8.11	15-29 min	Asa si asa	liceu	profesionala	Banala
4	4	8.47	30-59 min	Putin	postliceala	liceu	Obositoare
5	5	9.83	2-2.99 ore	Importanta	studii supe...	postliceala	Interesanta
6	6	9.36	3 ore sau...	Foarte im...	studii supe...	studii supe...	Incitanta
7	7	6.31	NR	NR	NR	NR	Inutila

În graficul 2.6 este prezentat rezultatul introducerii datelor pentru 7 respondenți, în ambele din cele două modalități mai sus amintite.

În fereastra din partea stângă nu s-a activat opțiunea (Wiew) „Value Labels" pe când în fereastra din partea dreaptă a fost selectată respectiva opțiune. Pentru primul respondent, numărul chestionarului este 1, media anului trecut este 9.42 iar pentru variabila V2 (timpul alocat studiului într-o zi obișnuită) valoarea efectiv introdusă în baza de date este 4 (având eticheta „1-1,99 ore"). Pentru variabila V3 (Importanța școlii pentru reușita în viață), valoarea efectiv introdusă în baza de date este 4 iar în fereastra din dreapta apare eticheta acestei valori, adică „importantă". Pentru același prim respondent studiile mamei și ale tatălui sunt cuantificate cu 4 respectiv 5 (în fereastra din stânga) ceea ce semnifică, conform etichetelor atribuite acestor valori, „postliceală", respectiv „studii superioare".

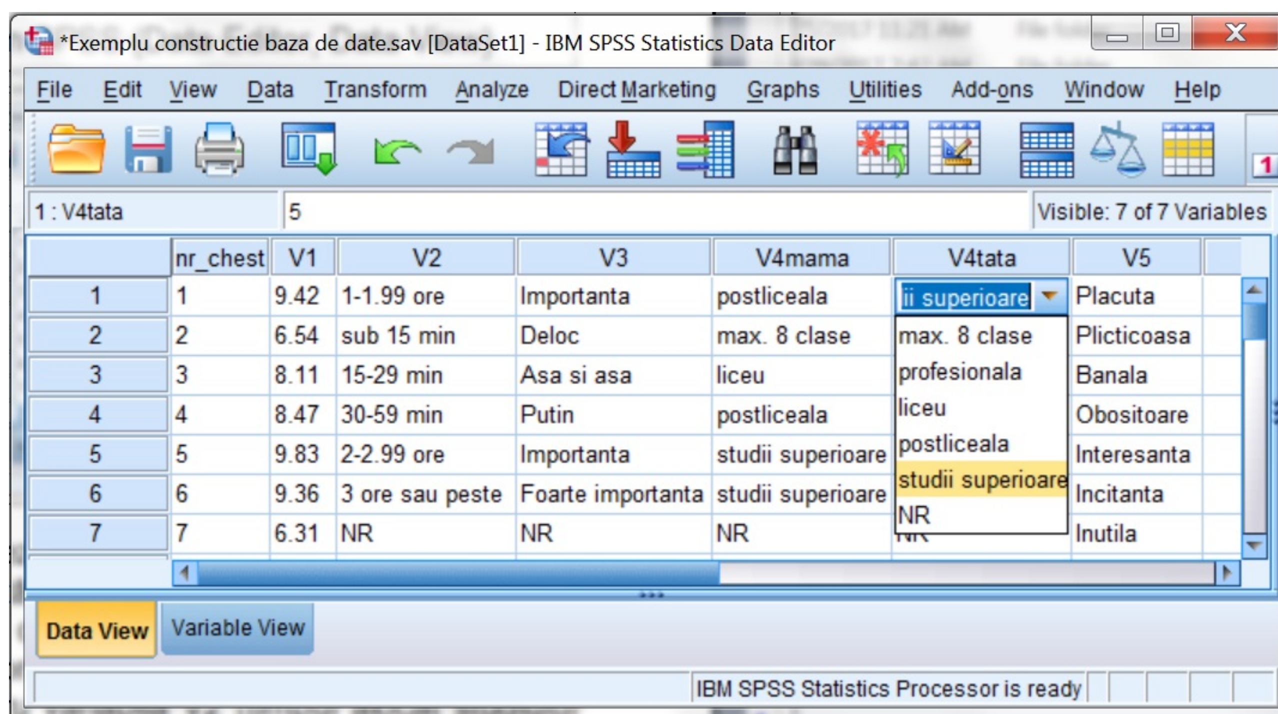
În cazul variabilei V5, desemnând modul în care respondentul ar caracteriza școala printr-un singur cuvânt, valorile efectiv introduse în baza de date (fereastra din stânga) sunt identice cu cele vizualizate în partea dreaptă deoarece în această situație nu s-au utilizat etichete pentru valori (în cazul unei întrebări deschise nu există, prin definiție, răspunsuri predeterminate). De altfel, și primele două variabile arată identic în cele două ferestre (care diferă doar prin activarea opțiunii „Value Labels") deoarece etichetele nu sunt necesare (numărul chestionarului, evident, este o valoare atribuită arbitrar, având ca scop să nu se preia de mai multe ori aceleași informații și să se poată verifica completitudinea introducerii datelor). Pe de altă parte, nici pentru variabila „V1", desemnând media anului trecut, etichetele valorilor nu au sens deoarece valorile conțin întreaga informație necesară interpretării lor.

Dorim să facem o precizare: în cazul chestionarului numărul șapte s-au inregistrat patru nonrăspunsuri, toate codate cu cifra 9, și le-au fost atribuite etichetele „NR". A se consulta „Introduction to SPSS (version 16), for Windows, Practical workbook".

Indiferent dacă selectăm sau nu opțiunea (Wiew) „Value Labels”, introducerea datelor va decurge simplu: putem să ne deplasăm în interiorul tabelului cu ajutorul tastelor funcționale și să introducem de la tastatură informațiile care sunt proprii fiecărui respondent pentru variabilele definite în baza de date (coloanele tabelului).

Dacă introducem datele despre respondenți selectând opțiunea (Wiew) „Value Labels”, mai avem o posibilitate de transfer a informațiilor. Dacă dăm dublu click pe o căsuță din tabel apoi selectăm butonul din partea stângă, se va deschide o listă cu etichetele valorilor variabilei din care se va alege cea corespunzătoare valorii pe care dorim s-o introducem. A se consulta lucrarea „Introduction to SPSS (version 16), for Windows Practical workbook”.

Graficul 2.7 Introducerea datelor în SPSS („Value Labels” selectat)



Dacă optăm să introducem datele selecând etichetele corespunzătoare din căsuțele tabelului, avantajul va fi că va scădea riscul erorilor (tastarea unor numere inadecvate valorilor variabilei). În graficul 2.7 se poate vedea că pentru primul respondent, în cazul variabilei „V4tata”, desemnând studiile tatălui, din lista răspunsurilor posibile a fost aleasă varianta „studii superioare” (prin click simplu). În mod analog se procedează pentru orice variabilă pentru ale cărei valori au fost definite etichete (cu ajutorul variantei de prezentare „data view”).

Introducerea datelor selectând „Wiew” „Value Labels” și alegând, în fiecare căsuță a tabelului (din Data Editor) eticheta valorii pe care dorim s-o introducem este, foarte probabil, o modalitate mai sigură de transfer a informațiilor dar necesită mai mult timp decât tastarea directă a datelor și deplasarea în tabel cu ajutorul tastelor funcționale.

Introducerea datelor în PSPP este foarte asemănătoare de cea realizată în SPSS iar dacă se optează pentru vizualizarea valorilor din baza de date, procedeele sunt identice (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1). Dacă se alege vizualizarea etichetelor valorilor din baza de date diferențele de procedură sunt minore.

În graficul 2.8 prezentăm modul de introducere a datelor atunci când este selectată opțiunea (Wiew) „Value Labels” deoarece aici am identificat o singură diferență: se va da click simplu într-una din căsuțele tabelului „Data Editor”, după care se va da click asupra săgeții din partea dreaptă a respectivei căsuțe. Va apărea lista cu etichetele variabilei din coloana corspunzătoare căsuței în care ne-am poziționat și vom putea alege valoarea pe care s-o atribuim răspunsului dat de respondentul ale cărui date sunt trecute pe linia pe care ne aflăm.

Graficul 2.8 Introducerea datelor în PSPP („Value Labels” selectat)

*Exemplu constructie baza de date.sav [DataSet1] — PSPPIRE Data Editor

File Edit View Data Transform Analyze Graphs Utilities Windows Help

1 : V4tata | studii superioare

Case	nr_chest	V1	V2	V3	V4mama	V4tata	V5
1	1	9.42	1-1.99 ore	Importanta	postliceala	superioare	Placuta
2	2	6.54	sub 15 min	Deloc	max. 8 clase	max. 8 clase	Plicticoasa
3	3	8.11	15-29 min	Asa si asa	liceu	profesionala	Banala
4	4	8.47	30-59 min	Putin	postliceala	liceu	Obositoare
5	5	9.83	2-2.99 ore	Importanta	studii superioare	postliceala	Interesanta
6	6	9.36	3 ore sau peste	Foarte importanta	studii superioare	studii superioare	Incitanta
7	7	6.31	NR	NR	NR	NR	Inutila

Data View Variable View

Filter off | Weights off | No Split

În graficul 2.8 se poate vedea că pentru variabila având denumirea „V4tata” și eticheta „Studiile tatălui”, în cazul chestionarului cu numărul 1, nivelul de școlaritate care va fi atribuit este „studii superioare” (Basic Stats in PSPP For NICAR 2016 Denver).

Facem o precizare, deși nu e legată direct de crearea bazelor de date și de introducerea informațiilor în structurile care permit preluarea acestora. În SPSS, instrucțiunea Edit Undo permite anularea unui număr nedefinit de pași parcurși în introducerea datelor pe când în PSPP această opțiune nu este accesibilă (de fapt, comanda „Undo” lipsește pentru orice fel de acțiune realizată cu programul PSPP). Evident, este o deficiență minoră, se poate reveni la căsuța unde s-a înregistrat eroarea și valoarea poate fi înlocuită prin tastarea celei corecte sau prin selectarea etichetei adecvate (din lista care apare în cazul în care introducem datele vizualizând etichetele acestora). În cazul altor proceduri ale programului PSPP, lipsa comenzii „Undo” poate crea probleme serioase (cum af fi situațiile în care sunt folosite comenzi de tipul „Cut”, „Copy”, „Paste”, ”Transform Compute” sau „Transform, recode into same variable”.

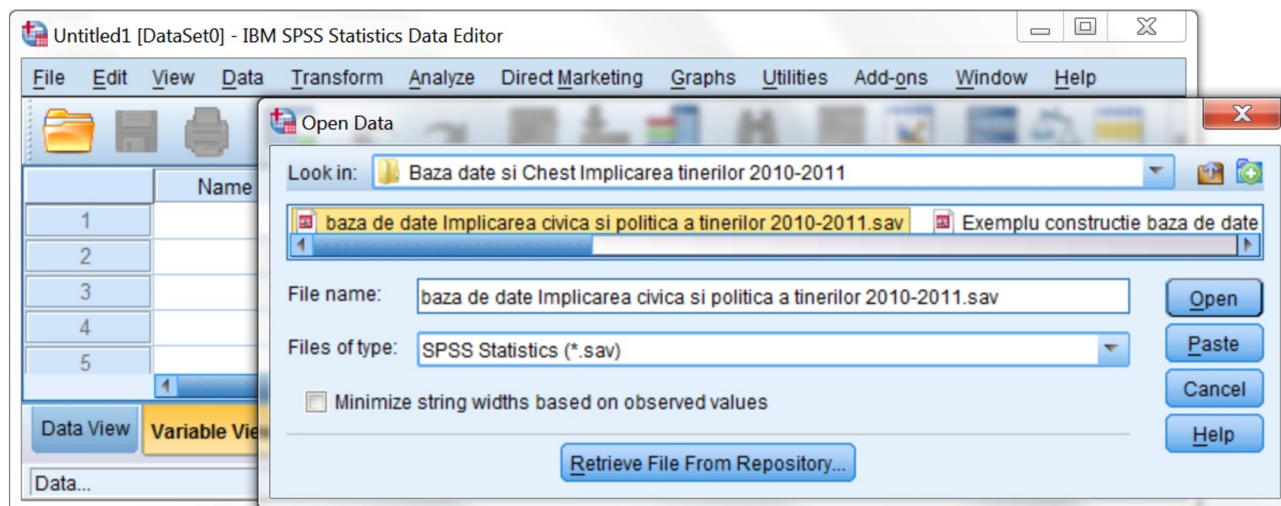
Mai menționăm faptul că pentru variabila string „V5” se poate, totuși, atașa o etichetă atât în SPSS cât și în PSPP: 9 semnifică nonrăspuns (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1)

3. CITIREA DATELOR ÎN SPSS ŞI PSPP

3.1 Citirea unui fişier tip bază de date (extensia .sav) SPSS şi PSPP

După cum am precizat deja, fereastra „Data Editor” utilizează baze de date cu extensia .sav. Dacă dăm comanda **File** şi alegem **Open** urmat de **Data**, vom obţine fereastra de dialog prezentată în graficul 3.1. (A Handbook of Statistical Analyses using SPSS).

Graficul 3.1 Deschiderea unui fisier tip bază de date (.sav) în SPSS



Pentru a indica unde se află fişierul tip bază de date SPSS (cu extensia .sav) pe care îl vizăm e necesar să arătăm calea înspre folderul care îl conţine prin intermediul căsuţei „Look in:”. În cazul nostru, baza de date pe care dorim s-o deschidem se află în folderul „Baza date si Chest Implicarea tinerilor 2010-2011” iar numele fisierului este „baza de date Implicarea civica si politica a tinerilor 2010-2011”. Menţionăm că în fereastra de dialog vor fi vizualizate fişierele având aceeaşi extensie cu cea care a fost setată în căsuţa „Files of type”. De notat că utilizând căsuţa de dialog „Files of type” putem afla toate tipurile de fişiere conţinând informaţii statistice pe care SPSS le poate interpreta sub forma bazelor de date şi converti la extensia „.sav”.

În graficul 3.2 apar informaţiile din baza de date care a fost deschisă (IBM SPSS Statistics 20 Core System, User’s Guide).

Graficul 3.2 Baza de date cu extensia „.sav” deschisă în SPSS

*baza de date Implicarea civica si politica a tinerilor 2010-2011.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

3 : codScoala 13135 Visible: 293 of 293 Variables

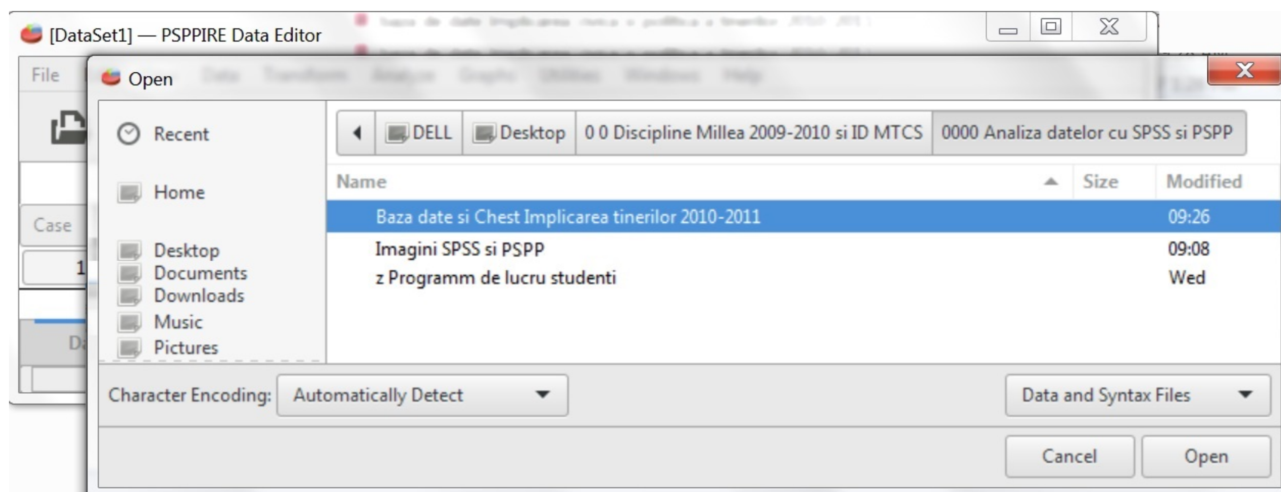
	judet	localitate	codScoala	clasa	scoala	a1_1	a1_2	a1_3
1	Botosani	Botosani	10103	a IX-a A	liceul teoretic n. iorga	Mai degraba de...	Cu totul de acord	Cu totul de acord
2	Suceava	Clg. Mold.	13135	a IX-a A	colegiul silvic bucovina	Mai degraba de...	Mai degraba de...	Mai degraba de...
3	Suceava	Clg. Mold.	13135	a IX-a A	colegiul silvic bucovina	Mai degraba de...	Cu totul de acord	Mai degraba de...

Data View Variable View

IBM SPSS Statistics Processor is ready

Deschiderea unei baze de date cu extensia „.sav” în programul PSPP este foarte asemănătoare cu procedura aplicată în SPSS (Basic Stats in PSPP For NICAR 2016 Denver).

Graficul 3.3 Deschiderea unui fisier tip bază de date (.sav) în PSPP

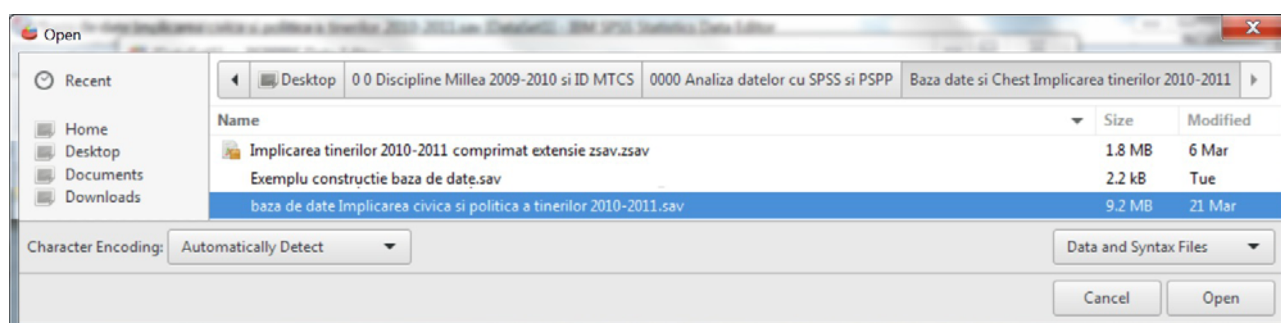


După ce dăm comenzile „File” și „Open”, se va deschide fereastra de dialog prezentată în graficul 3.3. Pentru a indica ce cale trebuie urmată pentru a ajunge la fișierul pe care dorim sa-l deschidem va trebui, în primul rând, să indicăm (din stânga ferestrei) punctul de plecare al căutării: „Desktop”, „Documents”, „Downloads”, „Pictures”, „Local Disk C”, etc. Succesiunea folderelor pe care le vom deschide pentru a ajunge la fisierul care ne interesează este vizibilă în partea de sus a ferestrei. Din întregul sistem (DELL) a fost ales, mai întâi, „Desktop-ul”, apoi de aici a fost selectat (cu click) „0 0 Discipline Millea 2009-2010 si ID MTCS” (care a fost „deschis” dând un nou „click). În continuare, din acest folder a fost selectat (prin click simplu) folderul numit „0000 Analiza datelor cu SPSS si PSPP”, al cărui conținut de foldere este vizualizat la mijlocul ferestrei de dialog din graficul 3.3:

- „Baza date si Chest Implicarea tinerilor 2010-2011”
- „Imagini SPSS si SPSS”
- „z Programm de lucru studenti”

Următorul pas constă în a selecta folderul care ne interesează („Baza date si Chest Implicarea tinerilor 2010-2011”) care conține baza de date pe care dorim s-o deschidem (marcată în zona mediană a graficului 3.4).

Graficul 3.4 Selectarea bazei de date (.sav) care va fi deschisă în PSPP



După ce am selectat fișierul tip bază de date denumit „baza de date Implicarea civica si politica a tinerilor 2010-2011”, dăm comanda „Open” (butonul din dreapta jos) și datele vor fi accesibile în „PSPP Data Editor”.

3.2 Citirea, cu SPSS, a datelor din Excel

Programul SPSS 20 permite citirea fișierelor create în Excel 2007 – 2010 și în versiunile anterioare. Există, însă, câteva cerințe / limitări legate de organizarea datelor în fișierele de tip Workbook: dacă dorim să poată fi transferată în SPSS o foaie de lucru „Excel” (cu extensia .xlsx), e important să tastăm în prima linie a coloanei - desemnând o anumită variabila statistică – numele respectivei variabile, fără spații între cuvinte (se poate folosi liniuța din partea de jos). Dacă se lasă spații între cuvinte, ele vor fi eliminate în momentul în care programul SPSS deschide foaia de lucru „Excel” (va atașa, totuși, o etichetă variabilei pentru care a eliminat spațiile libere din nume, constând chiar în numele care a apărut inițial în prima căsuță din coloana Excel). De notat că dintr-un fișier Excel (Workbook), conținând mai multe foi de lucru (Worksheets), va fi transferată în SPSS cea indicată într-o fereastră de dialog deschisă de programul SPSS.

Graficul 3.5 Foaia de lucru „Excel” pregătită pentru transferul în SPSS

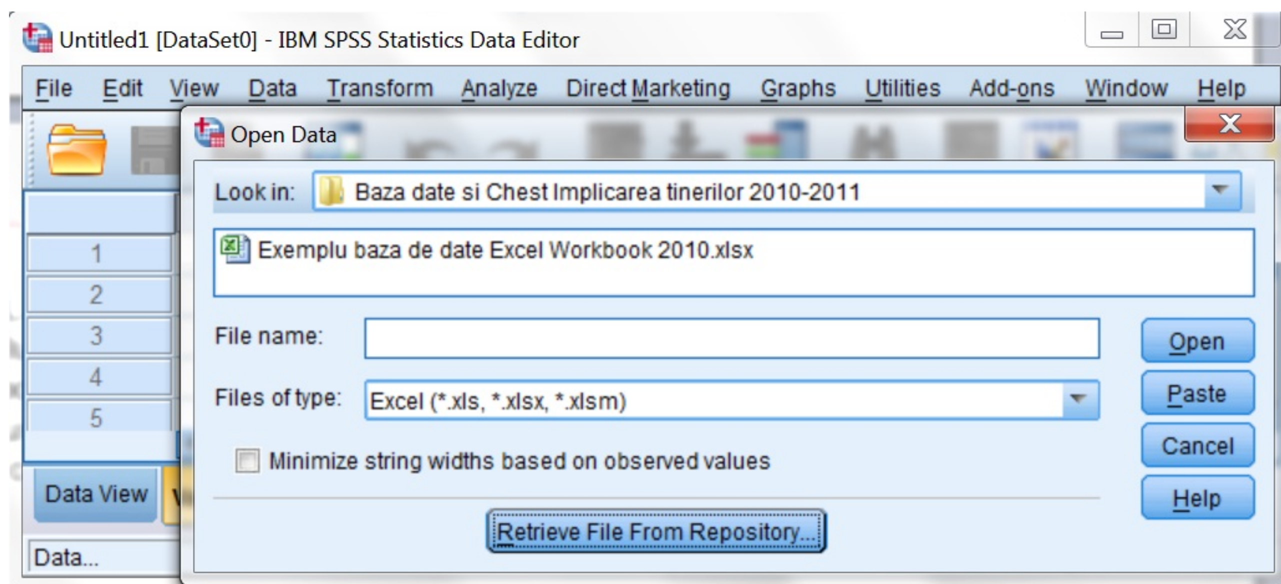
	A	B	C	D	E	F
1	Varsta	Sex	Studii	Domiciliu	Multumit_de_salariu	Interes pentru concedii in UE
2	18	1	2	Alba Iulia	1	2
3	22	0	3	Aiud	2	2
4	29	1	4	Galda de Sus	3	4
5	54	0	2	Ighiu	1	1
6	36	1	1	Stremt	2	1
7	67	1	5	Cugir	4	5
8	33	1	5	Baia de Aries	5	5
9	78	0	4	Teius	2	1

În foaia de lucru Excel (fișier cu extensia .xlsx) au fost introduse date pentru 6 variabile, cinci sub formă numerică (desemnând vârsta subiecților, sexul, nivelul studiilor și gradul de mulțumire față de salariu respectiv interesul manifestat pentru concedii în UE) și una alfanumerică (string) indicând localitatea de domiciliu.

După ce fișierul a fost salvat fără nici o modificare a tipului / extensiei acestuia, el va putea fi deschis din programul SPSS (IBM SPSS Statistics 20 Core System, User’s Guide).

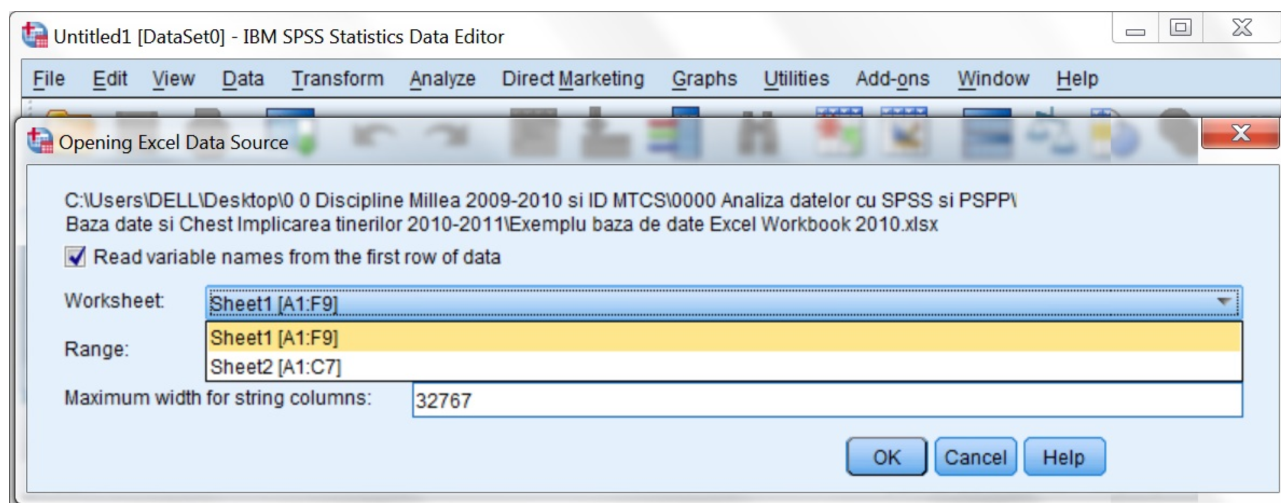
Lansăm programul SPSS şi apoi indicăm calea necesară deschiderii fişierului pe care l-am construit în „Excel”, numit „Exemplu baza de date Excel Workbook 2010”. Am păstrat titulatura „Workbook” deoarece, deşi am introdus date doar într-o singură foaie (Worksheet), nu am modificat structura de tip complex, conţinând mai multe „foi”, a fişierului iniţial.

Graficul 3.6 Deschiderea din SPSS a unei foi de lucru din fisierul Workbook



După ca a fost lansat SPSS, s-a dat comanda „File”, „Open” „Data” şi apoi, în fereastra de dialog prezentată în graficul 3.6 au fost indicate succesiv folderele ce trebuiau deschise pentru a se ajunge la fişierul denumit „Exemplu baza de date Excel Workbook 2010”. Pentrua fi vizualizate fişierele „Workbook” (cu extensia .xlsx) existente în folderul „Baza date si Chest Implicarea tinerilor 2010-2011” a fost selectat programul „Excel” din căsuţa cu titulatura „Files of type:”. Ultimul pas constă în selectarea fişierului excel care ne interesează şi iniţerea comenzii „Open”. După cum am arătat deja, se va deschide o fereastră de dialog care permite alegerea foii de lucru dorite - dintre cele ce conţin date (IBM SPSS Statistics 20 Core System, User’s Guide).

Graficul 3.7 Selectatea foii de lucru excel (Worksheet) deschisă în SPSS



Menționăm faptul că în partea de sus a ferestrei de dialog (graficul 3.7) este prezentată calea completă prin care se ajunge la fisierul „Excel” ce urmează a fi deschis. În aceeași fereasră este important să selectăm „Read variable names from the first row of data” pentru ca programul SPSS să atribuie fiecărei variabile numele care apare în prima căsuță a colanei din foaia de lucru „Excel” (IBM SPSS Statistics 20 Core System, User’s Guide).

În partea de jos a ferestrei de dialog (graficul 3.7) se poate stabili numărul de caractere acceptat pentru variabilele „String” (șir de caractere) pentru care programul SPSS impune o limită maximă de 32767). Facem precizarea că SPSS va interpreta ca numerice coloanele din „Excel” care vor conține doar numere sau căsuțe goale și va transforma în variabile „String” toate celelalte coloane din foaia de lucru „Excel”.

În graficul 3.7 apare și opțiunea „Range” care permite să se selecteze, din foaia de lucru „Excel”, care va fi porțiunea care va fi preluată în SPSS. În grafic apare „Sheet1[A1:F9]”, ceea ce înseamnă că se va prelua în SPSS prima foaie de lucru („Sheet1”) începând cu prima coloană și primul rând („A1”) și mergând până la a șasea coloană și la cea de a noua căsuță a acesteia („F9”). Folosind căsuța „Range” am fi putut opta, de exemplu, să eliminăm prima coloana, indicând „Range” B1:F9 (sau puteam elimina ultimul rând al foii de lucru, indicând „Range” A1:F8). Opțiunea descrisă anterior poate fi utilă în condițiile în care într-o registru de lucru Excel (Workbook) există informații complexe și suntem interesați să transferăm succesiv în SPSS informațiile existente în diferitele foi de lucru (Worksheet) sau chiar datele din interiorul acestora (Official IBM SPSS Statistics 21 Brief Guide).

Graficul 3.8 foaia de lucru deschisa cu SPSS

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

2 : Multumit_de_salariu 2 Visible: 6 of 6 Variables

	Varsta	Sex	Studii	Domiciliu	Multumit_de_salariu	InterespentruconcediiinUE	var	var
1	18	1	2	Alba Iulia	1			
2	22	0	3	Aiud	2	2		
3	29	1	4	Galda de Sus	3	4		
4	54	0	2	Ighiu	1	1		
5	36	1	1	Stremt	2	1		
6	67	1	5	Cugir	4	5		
7	33	1	5	Baia de Aries	5	5		
8	78	0	4	Teius	2	1		

Interes pentru concedii in UE

Data View Variable View

IBM SPSS Statistics Processor is ready

În graficul 3.8 e prezentată baza de date SPSS obținută prin transferul foii de lucru „Sheet1[A1:F9]”. De notat că pentru primele cinci variabile, în al căror nume nu s-au lăsat spații libere, programul SPSS nu a alocat etichete. Pentru cea de a șasea variabilă, căreia, în „Excel”, i s-a dat numele „Interes pentru concedii in UE”, programul SPSS a atribuit numele eliminând toate spațiile libere existente între cele 5 cuvinte („InterespentruconcediiinUE”) și a alocat o etichetă identică denumirii folosită inițial în „Excel”.

Este important să arătăm că toate observațiile privind transferul în SPSS a unui fisier „Excel Workbook 2010" sunt valabile și pentru formatul mai vechi, „Excel Workbook 97-2003" (SPSS 12.0 Brief Guide sau SPSS for Beginners Copyright 1999 Vijay Gupta).

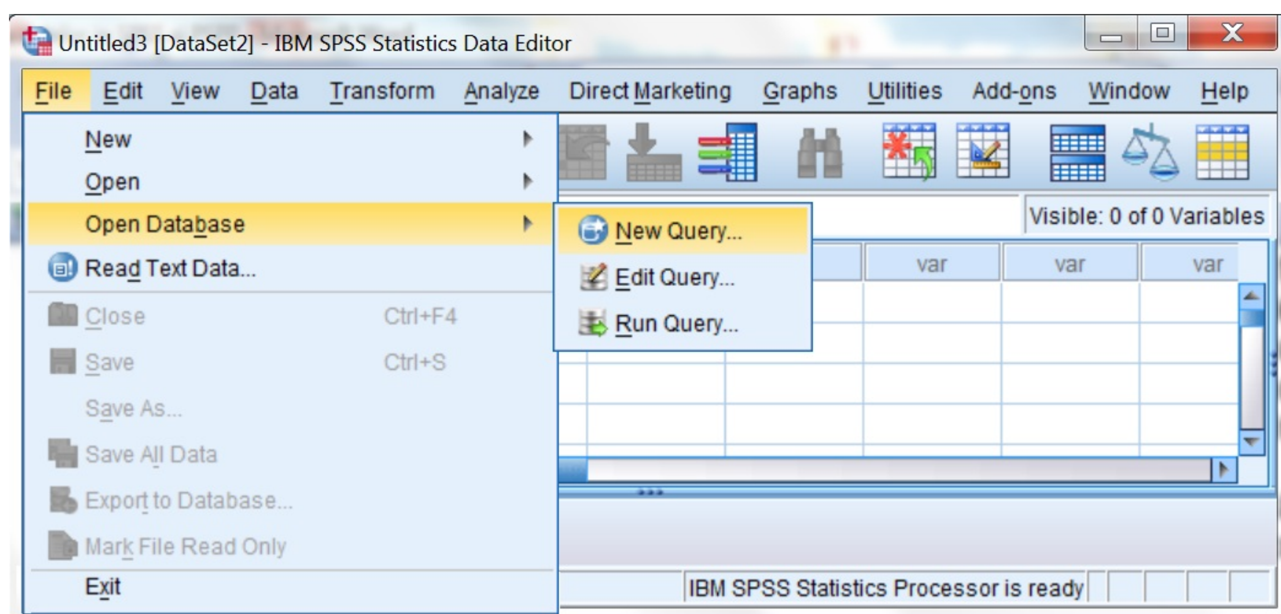
Cu SPSS 20 pot fi preluate și baze de date din foi de lucru („Worksheet Excel 4.0). Transferul către SPSS se va face în același mod ca și pentru „Excel Workbook 2010", cu diferența, evidentă, că nu se va mai utiliza opțiunea „Worksheet" pentru a selecta una dintre foile de lucru. O altă diferență, minoră, este că, în mod standard, variabilele numerice vor fi considerate ca având 2 zecimale dar în baza de date se vor păstra toate cifrele semnificative (de exemplu, dacă în „Excel Worksheet 4.0" s-au introdus 5 zecimale, în SPSS vor fi vizibile primele două dar înlocuind valoarea 2 cu valoarea 5 pentru numărul zecimalelor, acestea vor fi în întregime vizibile și se va putea opera cu ele.

Reamnntim faptul că PSPP nu dispune de faclități care să-i permită transferul informațiilor din Excel sau dintr-un program de gestionare a bazelor de date cum ar Access sau dBase, fapt care reprezintă o limitate importantă în raport cu SPSS.

3.3 Transferul, în SPSS, a datelor prin interogări ale bazelor de date

Pentru a avea acces la datele existente într-o bază de date (cum ar fi Access, dBase Microsoft Paradox, Microsoft Visual FoxPro, „Microsoft Access Text, Microsoft Oracle, SQL Server dar și pentru Excel) se va utiliza succesiunea de comenzi **File**, **Open Database**, **New Query**..., care va permite transferul global al informațiilor dar și selectarea variabilelor de interes sau a anumitor valori ale acestora (IBM SPSS Statistics 20 Core System, User's Guide).

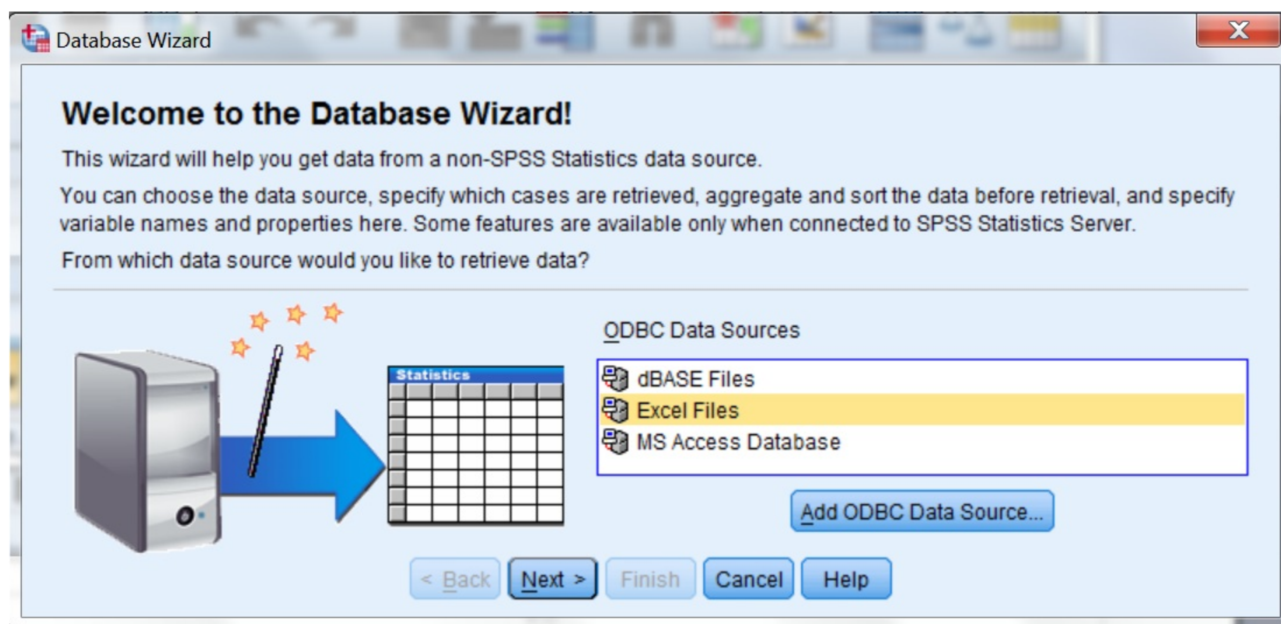
Graficul 3.9 Comenzile pentru transferul datelor în SPSS



Vom obține, în continuare, o ferastră de dialog care ne va permite să alegem programul din care dorim să transferăm datele înspre SPSS. Fereastra de dialog din SPSS 20 propune, implicit, soft-urile „dBASE", „Excel" și „Microsoft Access" ca posibile surse pentru baza de date care va fi construită în SPSS (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

De notat că SPSS 20 permite adăudarea altor surse de date care pot fi utilizate, cum ar fi „Microsoft Paradox” (cu extensia .db), „Microsoft Visual FoxPro”, „Microsoft Access Text” (.txt sau .csv), „Microsoft Oracle”, „SQL Server”, etc. Comanda este „ADD ODBC Data Source”(IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 3.10 Fereastra de dialog pentru selectarea sursei datelor

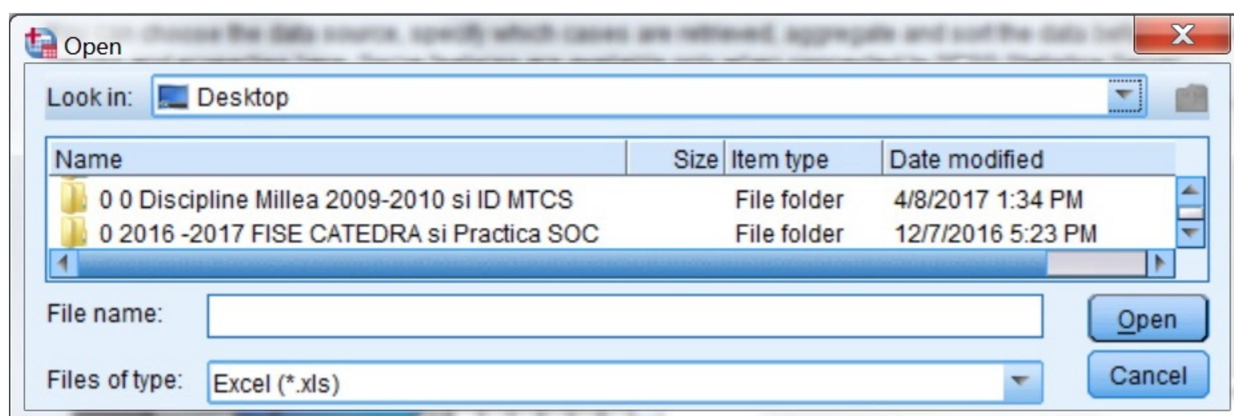


Vom începe prin a indica sursa datelor ca fiind Excel, în ideea de a releva diferențele care apar în raport cu procedura de accesare a unei sigure foi de lucru „Excel” (prin comanda „File”, „Open”, „Data”).

După cum suntem informați în fereastra care lansează „Database Wizard”, prin această procedură putem alege sursa exactă a datelor, putem specifica ce cazuri dorim să selectăm din respectivele surse de date, avem posibilitatea de a agrega și ordona bazele de date pe care le preluăm și putem, de asemenea, să modificăm / precizăm numele și tipul unora dintre variabile.

După ce am ales „ODBC Data Sources” vom da click pe butonul „next” și se va deschide o fereastră de dialog care permite să indicăm fisierul Excel 97-2003 (cu extensia .xls, nu .xlsl) pe care dorim sa-l transferăm în SPSS (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

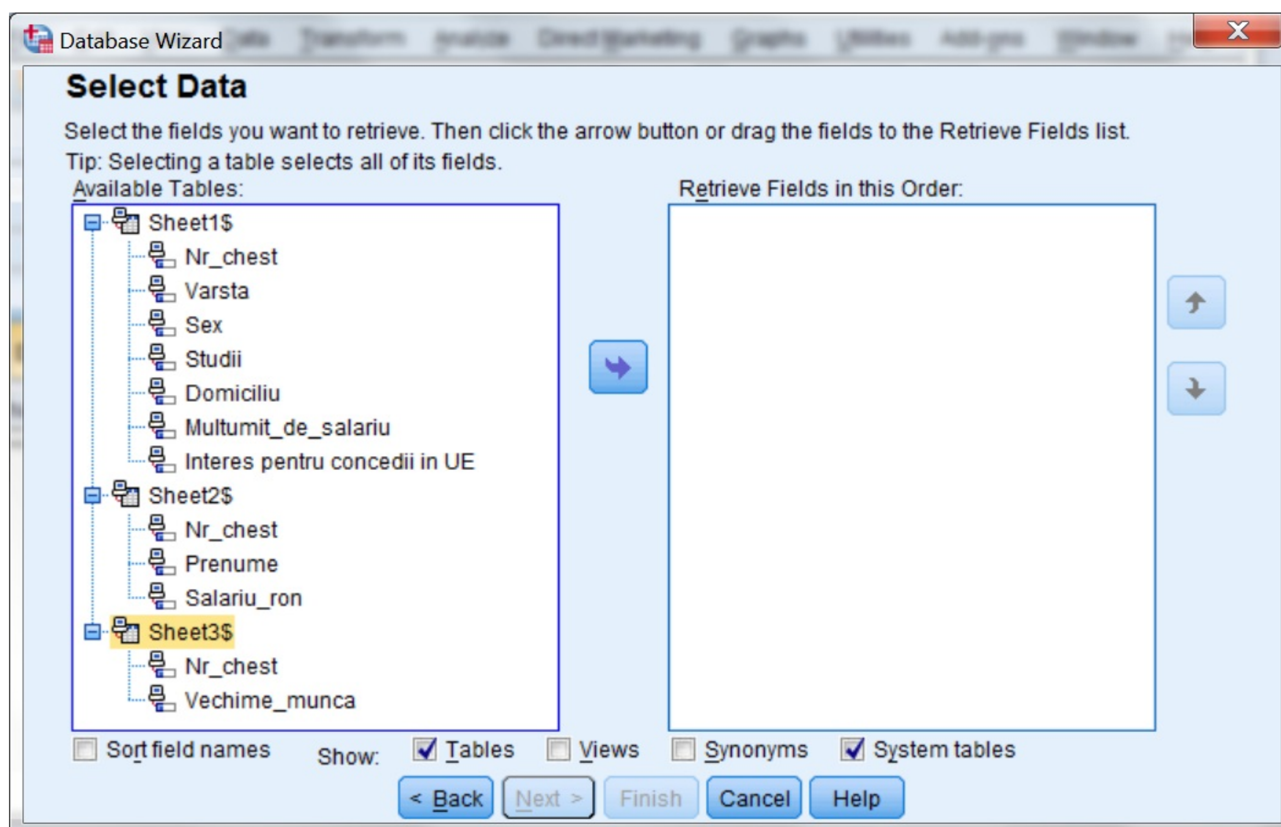
Graficul 3.11 Indicarea căii de acces spre fisierul Excel (Workbook)



Pașii succesivi care trebuie urmați pentru a ajunge la folderul care conține fișierul de interes pentru noi („Exemplu baza de date Excel Workbook 97-2003”) vor fi indicați din căsuța „Look in:” .Când numele folderului căutat va apărea în căsuța „Look in:”, fisierele Excel (cu extensia .xls, selectată din căsuța „Files of type:”) conținute de acesta vor fi vizualizate în zona centrală a feresterei (cu titulatura „File Name”, graficul 3.11). Vom selecta fisierul care ne interesează dând click simplu și apoi vom da comanda de deschidere a fișierului de la butonul „Open” (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

În graficul 3.12 sunt prezentate cele trei foi de lucru (woksheet) care sunt conținute în fișierul („Workbook”) din care vor fi selectate datele, fiind vizualizate toate variabilele din respectivele foi de lucru.

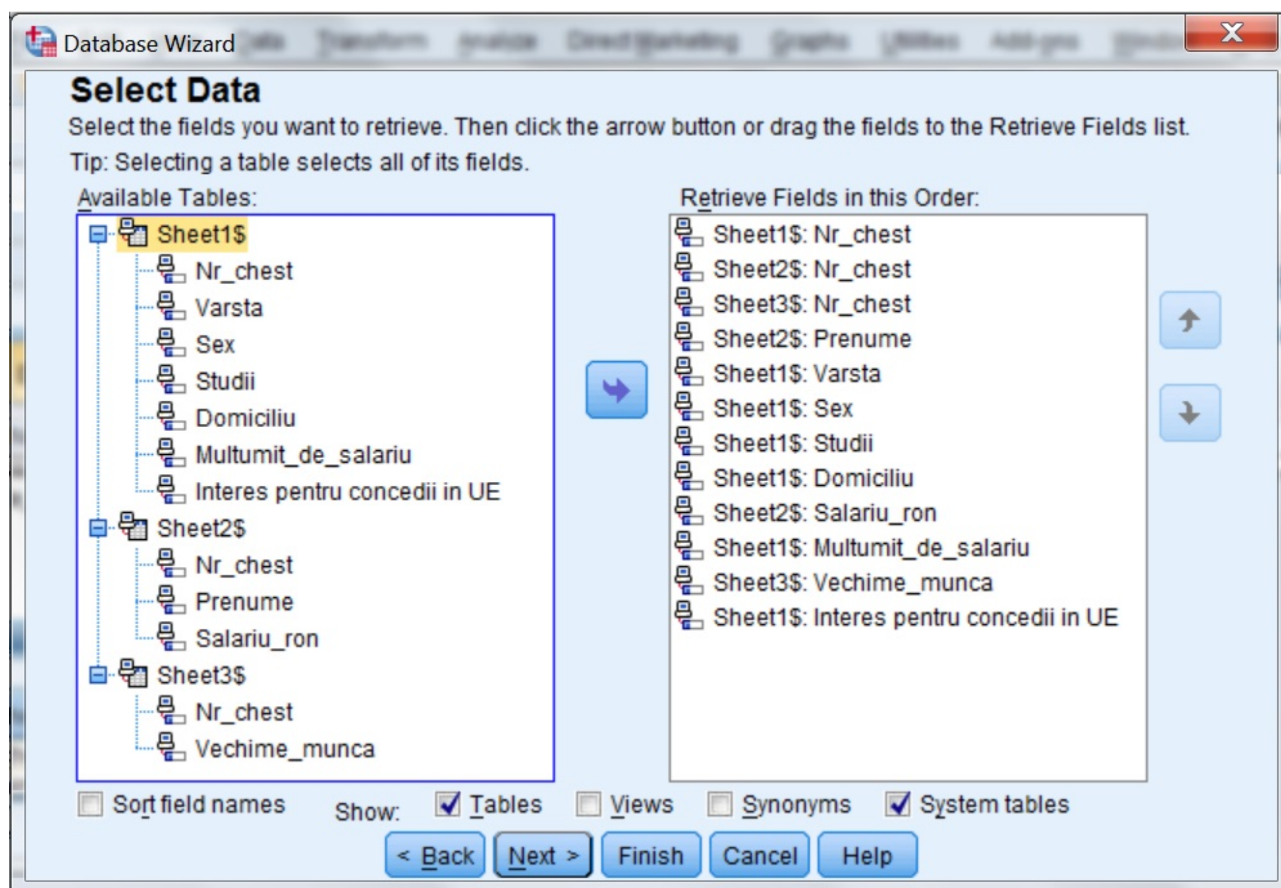
Graficul 3.12 Variabilele ce pot fi agregate din foile de lucru Workbook Excel



Variabilele pot fi selectate (toate sau doar unele dintre ele) și transmise în partea dreaptă a ferestrei în ce ordine dorim. Procedura pe care am lansat-o ne va permite, în final, să construin o bază de date SPSS în care sunt agregate datele variabilelor pe care le-am selectat, în ordinea pe care am stabilit-o.

O astfel de agregare are sens dacă variabilele din baza de date finală se vor referi, pe o linie, la aceleași unități statistice. Ca atare, este necesar să existe o variabilă care să facă legătura între foile de lucru cu care operăm. Am putea să ne gândim că această variabilă este numărul chestionarului, dacă în fiecare foaie de lucru s-au introdus date dintr-o pagină din cele 3 ale unui chestionar (sau am putea apela la CNP-ul respondenților, care ar trebui să fie prezent în fiecare din cele trei foi de lucru ale fișierului „Excel Workbook”. Agregarea se va realiza cu ajutorul variabilelor indicate ca făcând legătura dintre cele trei foi de lucru. (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Graficul 3.13 Variabilele selectate pentru agregarea datelor Workbook Excel



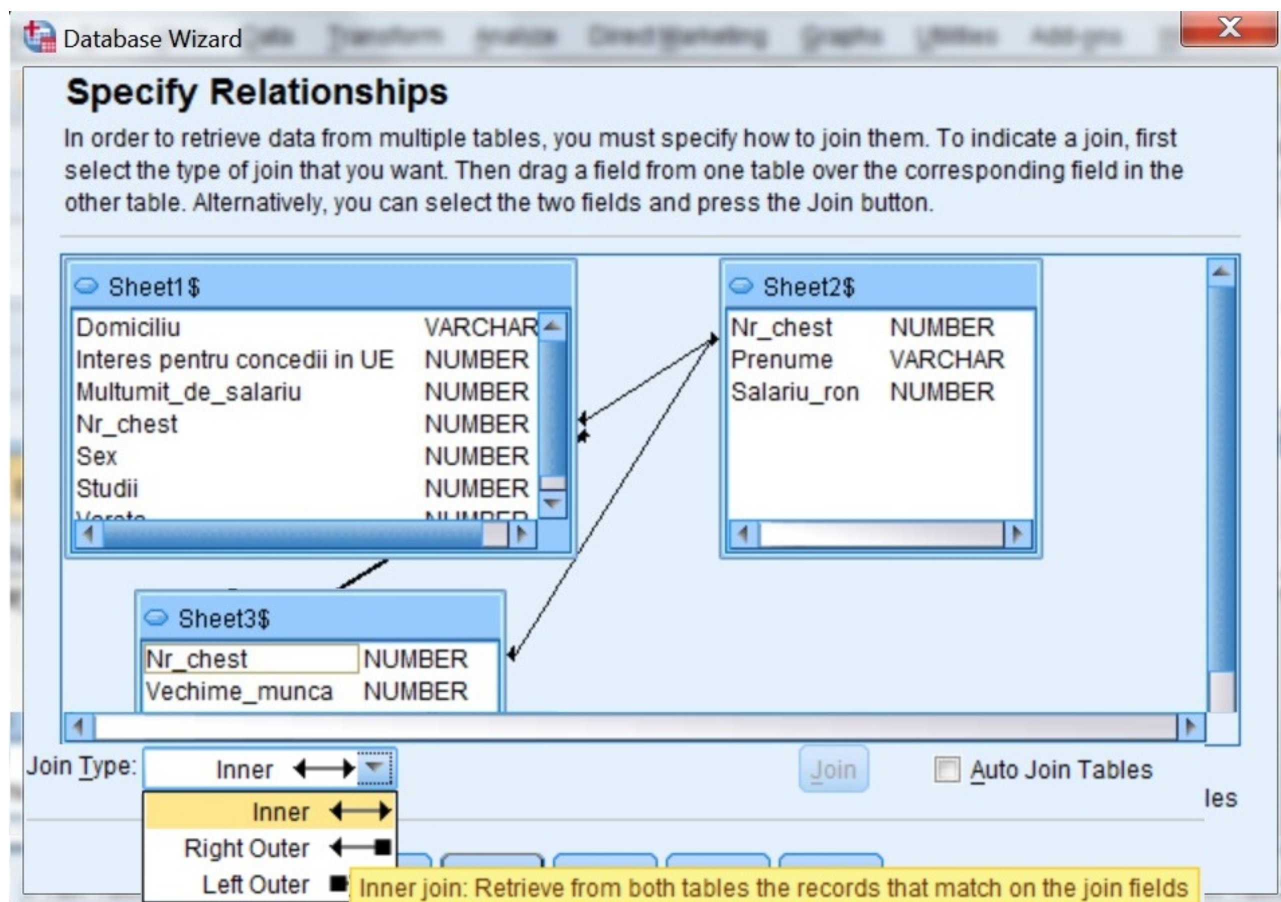
În dreapta graficului 3.13 („Retrieve fields in this Order”) primele 3 variabile poartă același nume, ca să releve faptul că vor fi utilizate pentru a face legătura dintre unitățile statistice despre care există informații în foile de lucru „Excel”. De notat faptul că în situația cea mai simplă, fiecare dintre cele 3 variabile are același număr de valori, fiecare valoare fiind unică pentru o anumită unitate statistică. În consecință, fiecare „worksheet” din cele 3 poate fi perfect ordonată crescător sau descrescător iar numărul înregistrărilor este același (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Ordinea variabilelor din partea dreaptă a graficului 3.13 va fi cea pe care o vom regăsi, în final, în baza de date construită în programul SPSS prin agregarea celor 3 foi de lucru Excel. Deoarece programul SPSS nu acceptă caracteristici cu același nume, pentru variabilele cu numele „nr_chest” se vor adăuga cifre care să diferențieze caracteristicile între ele (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

După selectatea variabilelor și a ordinii acestora vom da comanda „Next” din butonul aflat la baza ferestrei „Select data” (a se vedea graficul 3.13). Vom obține, astfel, o nouă fereastră de dialog, „Specify Relationships” (graficul 3.14) care ne va permite să indicăm modalitatea de a realiza agregarea corectă între cele 3 foi de lucru „Excel”.

Dacă se va selecta opțiunea „Auto Join Tables” (partea de jos a graficului 3.14), programul SPSS va identifica automat variabila având același nume în cele 3 foi de lucru. În caz contrar, va trebui să se specifice, pentru fiecare pereche de „worksheet”, care sunt cele două variabile care vor face legătura dintre ele prin selectare – click simplu – și comanda „Join” (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 3.14 Selecția variabilelor care permit legarea foilor de lucru Excel



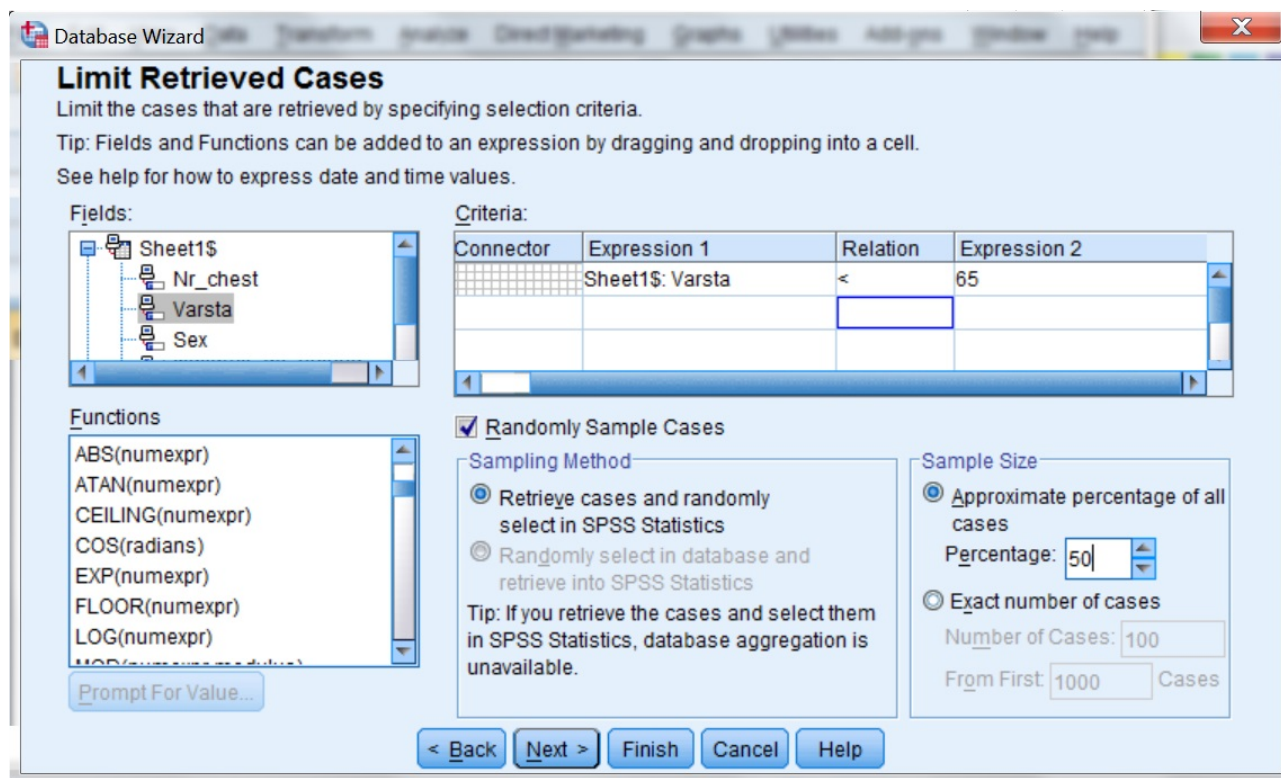
În cazul în care există trei sau mai multe foi de lucru excel, butonul „Joint Type” nu este operant: singura variantă posibilă este cea cu titulatura „Inner”. În acest caz, dintre valorile celor trei variabile care au fost definite ca făcând legătura între foile de lucru „Excel”, vor fi păstrate doar acele care se regăsesc în fiecare dintre respectivele foi. În consecință, cazurile (liniile) din baza de date finală din SPSS vor fi cele pentru care s-au identificat aceleași valori ale celor 3 variabile definite în relații bilaterale prin „Joint”. Restul cazurilor din foile de lucru - cu valori neconcordante - vor fi eliminate. (IBM SPSS Statistics 20 Core System, User's Guide, 2011)

În cazul în care s-ar fi optat pentru agregarea a numai două foi de lucru „Excel” exista posibilitatea alegerii unui „Worksheet” pentru care să fie păstrate și cazurile (liniile) care au valori ale variabilei de legătură fără corespondent la nivelul celeilalte variabile indicată pentru agregarea celor două structuri de date. De exemplu, putem adăuga cazuri în Sheet1$ cu valori pentru „Nr_chest” între 100 și 106 (cu increment de o unitate) iar pentru Sheet2$ pot fi introduse, în plus, cazuri între 10 și 17 (același increment).

Dacă am selecta „Nr_chest” ca variabila de legătură dar pentru „Join” și, în plus, am opta pentru varianta „Right Outer”, s-ar păstra toate cazurile din „Sheet2$” (aflată în partea di dreapta a ferestrei de dialog) și doar cazurile comune celor două foi de lucru pentru „Sheet1$” (aflată în partea stângă a ferestrei). În mod analog, optând pentru „Left Outer” se păstrează toate cazurile pentru „Sheet1$” iar pentru „Sheet2$” rămân doar cazurile comune (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

După ca a fost aleasă varianta de de agregare a foilor de lucru „Excel” se va da comanda „Next”, în urma căreia se va deschide fereastra de dialog „Limit Retrived Cases” care permite să indicăm proceduri prin care pot fi limitate cazurile care vor fi selectate în baza de date finală din programul SPSS.

Graficul 3.15 Definirea regulilor de selectare (restricționare) a cazurilor



În graficul 3.15 s-a formulat cerința ca variabila „Varsta” din „Sheet1$” sa fie mai mică decât valoarea 65.

Pentru căsuțele „Experssion 1” și „Expression 2”, avem posibilitatea, dând click deasupra celei dorite, să alegem oricare dintre variabilele prezente în lista celor care au fost selectate pentru agregare („Fields”:).

Este posibil, de asemenea, ca în căsuțele mai sus amintite să introducem o valoare numerică sau o funcție care să conțină valori numerice sau nume ale variabilelor care au fost selectate. Pentru a folosi „Functions” trebuie să selectăm funcția dorită din partea stânga-jos a ferestrei și să o tragem cu mouse-ul deasupra căsuței din dreptul coloanei „Experssion 1” sau „Expression 2”. Expresia numerică pentru funcția aleasă se introduce de la tastatură și poate conține variabile selectate pentru agregare, numere sau simboluri aritmetice.

E important de știut că în coloana „Relation” putem selecta simbolurile „mai mic (sau egal)”, „egal”, „diferit” respectiv „mai mare (sau egal)”.

Când se folosesc mai multe condiționări pentru selectarea cazurilor ce vor fi vizate pentru agregare, se utilizează coloana „Connector” unde se va specifica dacă noua condiționare trebuie să fie simultană cu cele precedente (folosind conectorul logic „AND”) sau este doar alternativă acestora (folosind conectorul „OR”). De notat că pentru operatorii logici „și” respectiv „sau” e importantă ordinea operațiilor atunci când sunt combinați (PvQ este ehivalent cu QvP respectiv P&Q e identic cu Q&P dar pentru expresia PvQ&R contează ordinea aplicării operatorilor logici).

În tabelul 3.1 sunt prezentate câteva proprietăți importante ale operatorilor logici „și” („And”, în engleză, notat „&”) respectiv „sau” („Or”, în engleză, notat „V”).

Tabelul 3.1 Asociativitatea și distributivitatea conjuncției și disuncției

Asociativitatea conjuncţiei	[(p & q) & r] ≡ [p & (q & r)]
Asociativitatea disjuncţiei	[(p V q) V r] ≡ [p V (q V r)]
Distributivitatea conjuncţiei faţă de disjuncţie	[p & (q V r)] ≡ [(p & q) V (p & r)]
	[(p V q) & r] ≡ [(p & r) V (q & r)]
Distributivitatea disjuncţiei faţă de conjuncţie	[p V (q & r)] ≡ [(p V q) & (p V r)]
	[(p & q) V r] ≡ [(p V r) & (q V r)]

Având în vedere că prin opțiunea „Help” a programului SPSS 20 nu se obțin informații despre modul în care programul agregă mai multe cereri pentru selectarea cazurilor care vor intra în componența bazei finale de date din SPSS, sugerăm să nu se folosească succesiuni de propoziții logice pentru care ordinea operațiilor este importantă. În cazul a două propoziții se pote folosi, evident, „p V q” respectiv „p & q”, deorece există un singur mod de a stabili valoarea de adevăr a expresiilor. De asemenea, se pot utiliza succesiuni de conjuncții cu oricât de multe propoziții logice, expresia fiind adevărată dacă toate sunt adevărate și falsă dacă cel puțin una e falsă. Analog, sunt acceptate succesiuni de disjuncții fără limitări deoarece expresia e adevărată dacă cel puțin o propoziție e adevărată și falsă dacă toate propozițiile sunt false.

În cazul în care se fac combinații de tipul „p V q & r” respectiv „p & q V r” trebuie verificat în prealabil modul (ordinea) în care SPSS efectuează operațiile logice deoarece selectarea cazurilor depinde de modul în care au fost poziționate parantezele (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Un ultim aspect important în fereastra de dialog din graficul 3.15: căsuța „Randomly Sample Cases” oferă posibilitatea selectării unui eșantion aleator dintre cazurile care au fost alese pentru construirea bazei de date din SPSS.

O primă variantă este aceea de a prelua toate cazurile selectate în SPSS iar construirea eșantionului să se realizeze în cadrul acestui program („Retrive cases and randomly select in SPSS Statistics”). În acest caz s-ar putea lucra cu baze de date de foarte mari dimensiuni ceea ce ar mări timpul de lucru. Cealaltă variantă constă în a selecta aleator eșantionul pe măsură ce sunt extrase cazurile din bazele de date folosite ca sursă pentru programul SPSS („Randomly select in database and retrive into SPSS Statistics”).

Mărimea eșantionului poate fi aleasă cu o anumită aproximație, indicând ponderea acestuia din totalul cazurilor selectate pentru construcța bazei de date agregate în SPSS („Approximate percentage of all cases”) sau poate fi precizată în mod exact („Exact number of cases”). În cea de a doua situație va trebui să indicăm atât numărul de cazuri dorite în eșantion cât și numărul unităților statistice luate în calcul pentru selectarea acestora. Evident că este important să estimăm corect numărul total de cazuri disponibile în baza / bazele de date sursă (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

În graficul 3.15 am ales să selectăm cazurile pentru care variabila vârstă ia valori mai mici decât 65 (selecția a avut, evident, un caracter arbitrar, cu scop didactic). În plus, am ales să extragem un eșantion aleator cu un volum de aproximativ 50% din totalul datelor care respectă cerințele formulate până în acest moment. Menționăm că am ales ca procedura de eșantionare să se realizeze în programul SPSS, nu pe măsură ce sunt selectate cazurile din foile de lucru „Excel" („retrive cases and randomly selected in SPSS Statistics").

Dacă în fereastra prezentată în graficul 3.15 acționăm butonul „Next" obținem o ultimă modalitate de intervenție asupra variabilelor care vor constitui baza de date din SPSS. Astfel, se pot modifica numele variabilelor iar pentru cele care au fost definite ca șir de caractere („string") se poate opta pentru transformarea acestora în variabile numerice. Astfel, fiecărei valori numerice îi va corespunde valoarea alfanumerică („string"), în ordine alfabetică (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 3.16 Editarea finală a variabilelor ce vor fi agregte în baza SPSS

Database Wizard

Define Variables

Edit variable names and properties.
Names will be automatically generated if left blank.
To convert a string variable to numeric using the original values as value labels, check Recode to Numeric.

	Result Variable Name	Data Type	Recode to Numeric
Sheet1$: Nr_chest	Nr_chest	Numeric	
Sheet2$: Nr_chest	Nr_chest1	Numeric	
Sheet3$: Nr_chest	Nr_chest2	Numeric	
Sheet2$: Prenume	Prenume	String	☐
Sheet1$: Varsta	Varsta	Numeric	
Sheet1$: Sex	Sex	Numeric	
Sheet1$: Studii	Studii	Numeric	
Sheet1$: Domiciliu	Domiciliu	String	☐
Sheet2$: Salariu_ron	Salariu_ron	Numeric	
Sheet1$: Multumit_de_salariu	Multumit_de_salariu	Numeric	
Sheet3$: Vechime_munca	Vechime_munca	Numeric	
Sheet1$: Interes pentru concedii in UE	Interes_pentru_concedii_in_UE	Numeric	

Width for variable-width string fields: 255 ☑ Minimize string widths based on observed values

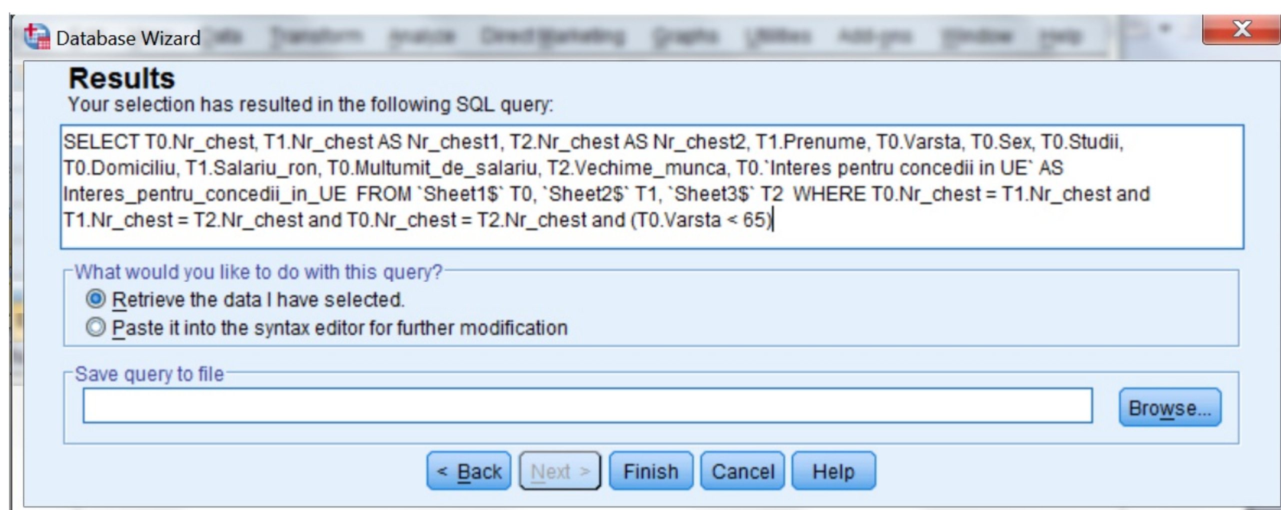
< Back | Next > | Finish | Cancel | Help

Fereastra de dialog mai oferă o opțiune, aceea de a stabili numărul maxim de caractere alocate variabielor alfanumerice. Din căsuța de dialog „Width for variable–width string fields" s-a specificat valoarea 255 în cazul concret prezentat în graficul 3.16). În aceeași fereastră se poate opta pentru minimizarea numărului de caractere alocat fiecărei variabile definită ca șir de caractere („Minimize string widths based on observed values") în funcție de lungimea efectivă a cuvintelor din care e alcătuită. De exemplu, dacă pentru valorile variabilei „Prenume" numărul maxim de caractere folosite pentru prenumele unui elev a fost de 64, atunci numărul de caractere alocat respectivei caracteristici va fi 64, nu 255 (a se vedea „IBM SPSS Statistics 20 Core System, User's Guide, 2011").

Graficul 3.17 oferă un rezumat al selectării datelor din programul SPSS prin apelarea foilor de lucru „Excel”. Sunt prezentate variabilele selectate în ordinea stabilită (indicându-se de fiecare dată foaia de lucru din care provine variabila). În fereastră sunt precizate și modificările aplicate denumirilor variabilelor, în primul rând pentru cele care fac legătura dintre foile de lucru „Excel”: numele „Nr_chest”, existent în foile de lucru notate T1 și T2, devine „Nr_chest1” și „Nr_chest2” în baza de date SPSS.

În plus, numele variabilei „interes pentru concedii în UE” devine interes_pentru_concedii_în_UE” având în vedere că în SPSS nu se acceptă spații în numele variabilelor. În continuare, rezultatul selecției variabilelor (descris prin „SQL query”) indică modul în care s-a realizat legătura dintre cele 3 foi „Excel” folosind variabilele cu numele „Nr_chest”. În final, este indicată condiția de selectare a cazurilor care a fost aplicată, adică „Varsta” <65 (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 3.17 Forma finală a interogării foilor de lucru „Excel” din SPSS



În final, dacă în fereastra de dialog alegem varianta „Retrive the data I have selected” și dăm clik pe butonul „Finish”, vom obține baza de date în SPSS în concordanță cu cerințele / specificările pe care le-am formulat IBM SPSS Statistics 20 Core System, User’s Guide, 2011.

Graficul 3.18 Baza de date din SPSS conform cerințelor formulate

*Untitled8 [DataSet7] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

1 : Nr_chest 1.00 Visible: 12 of 12 Variables

	Nr_chest	Nr_chest1	Nr_chest2	Prenume	Varsta	Sex	Studii	Domiciliu	Salariu_ron	Multumit_de_salariu	Vechime_munca	Interes_pentru_concedii_in_UE
1	1.00	1.00	1.00	Vasile	18.00	1.00	2.00	Alba Iulia	1000...	1.00	1.00	2.00
2	2.00	2.00	2.00	Dan	22.00	1.00	3.00	Aiud	2000...	2.00	3.00	2.00
3	7.00	7.00	7.00	Alexandru	33.00	1.00	5.00	Baia de Aries	2900...	5.00	11.00	5.00
4	5.00	5.00	5.00	Ion	36.00	1.00	1.00	Stremt	3000...	2.00	10.00	1.00
5	8.00	8.00	8.00	Marius	58.00	1.00	4.00	Teius	3300...	2.00	35.00	1.00

Data View Variable View

Mai menționăm faptul că interogarea foilor de lucru „Excel” pe care am formulat-o putea să fie salvată într-un fișier tip sintaxă (având extensia .sps) care va permite realizarea unor modificări ulterioare ale comenzilor care au fost formulate (și, evident, va permite controlul punerii în aplicare a respectivelor comenzi).

Dacă baza de date interogată din SPSS este cea proprie programului „Access”, pașii sunt aceiași dar trebuie să se țină seama de complexitatea respectivului mediu de gestionare a datelor.

În folderul „Samples” din „Office11”, situat în folderul „Microsoft Office”, sub-folderul „Program Files” există o bază de date „Access” care conține mai multe tabele, sub denumirea „Northwind” (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Graficul 3.19 Database cu tabele multiple „Access” (exemplu de structurare)

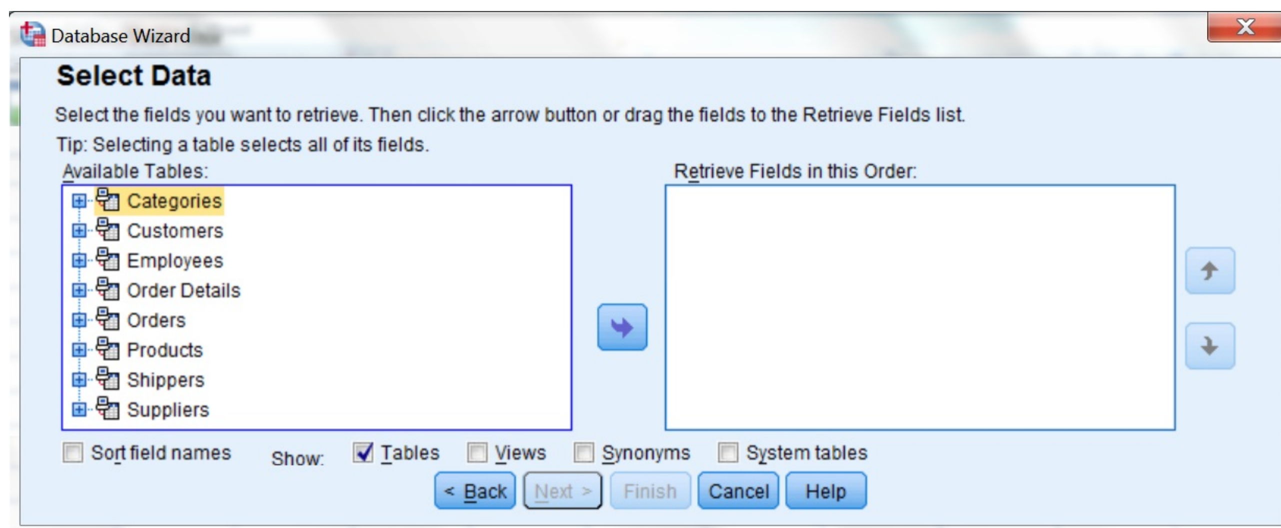
În graficul 19 este prezentată o bază de date construită în „Access” în care există opt tabele distincte („Categories”, „Customers”, „Employees”, etc.), fiecare tabel fiind structurat complex în funcție de anumite caracteristici. Am ales pentru exemplificare tabelul „Customers” pentru care, în tabelul de bază, primele variabile definite sunt „Customer”, conținând un cod restrâns al numelui clientului și apoi „Company Name” (numele firmei cu care se colaborează). Apar, în continuare, numele persoanei de contact („contact name”, poziția pe care o deține în firmă („Contact Title”), adresa („Adress”), Orașul („City”) și regiunea în care se află respectiva firmă.

La nivelul fiecărei firme clientă din baza de date se poate deschide un nou tabel conținând totalitatea livrărilor care au fost făcute înspre aceasta. Tabelul indică data prevăzută pentru livrare („Requred Date”), data efectivă la care s-a realizat livrarea („Shipped date”), calea de livrare („Ship Via”), cheltuielile de transport („Freight”), firma de destinație („Ship name”), adresa unde se face livrarea („Ship address”), orașul de destinație („City”), regiunea („Ship Region”) și codul poștal (Ship Postal).

În plus, pentru fiecare livrare se poate deschide un nou tabel care indică produsele care au fost livrate, prețul unitar, cantitatea și discauntul aplicat (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Dacă vom relua pașii din graficele 3.9, 3.10 şi 3.11 urmărind, de această dată, interogarea unei baze de date „Access”, în cazul graficului 3.11 vom alege „MS Access Database”. Pentru exemplul de bază de date definită în „Access” pe care l-am prezentat anterior, vom obține (după indicarea căii de acces spre fişierul „Northwind”) o fereastră de dialog similară cu cea existentă în graficul 3.13, care ne va permite să selectăm variabilele ce vor intra în componența bazei de date SPSS. Pentru mai multe detalii, a se consulta ghidul „IBM SPSS Statistics 20 Core System, User's Guide, 2011”.

Graficul 3.20 Tabelele care pot fi importate din „Access”



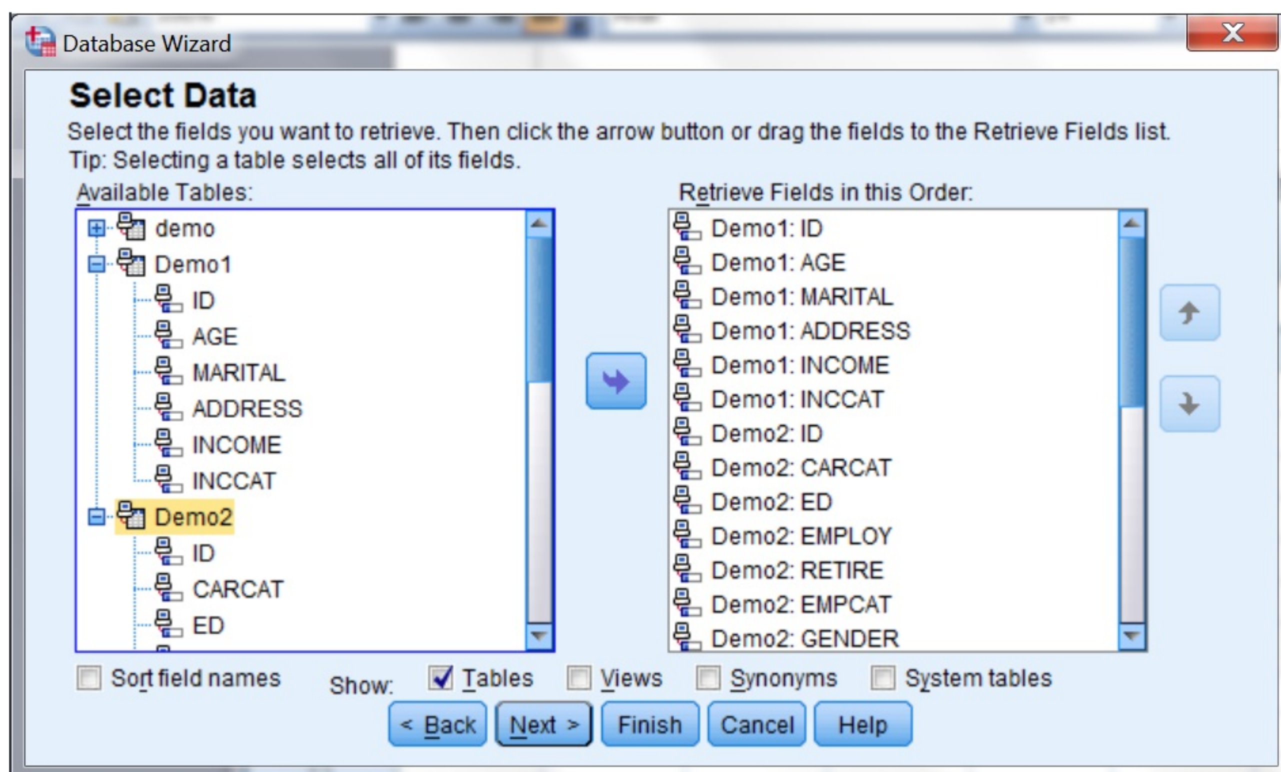
În Graficul 3.20 s-a optat doar pentru vizualizarea tabelelor definite în fisierul „Access” (opțiunea „Show: Tables”). În acest caz se vor putea importa în SPSS variabile selectate din unul sau mai multe dintre respectivele tabele. Dacă se vor selecta caracteristici din mai multe tabele, va fi necesar să se indice variabila (sau variabilele) care realizează legătura dintre respectivele tabele, în mod analog situației care a fost prezentată în graficul 3.14 (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Având în vedere că baza de date din „Access” este mult mai complex definită decât cea din „Excel”, există o serie de opțiuni SPSS care permit vizualizarea diferitelor aspecte structurale ale acesteia şi, implicit, importarea datelor sub acea formă. Opțiunea „Show” din partea de jos a ferestrei de dialog (graficul 3.20) permite să alegem variantele „Views” şi „System tables” care ne vor da posibilitatea să accesăm datele în multitudinea de variante în care au fost definite (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

În programul SPSS 20 există un exemplu de fişier „Access” numit „Demo” mult mai simplu (C / Progrm files, IBM, SPSS, Statistics, 20, Samples, English). În acest caz avem de-a face cu un singur tabel care poate fi peluat integral sau pot fi selectate doar unele dintre variabilele acestuia.

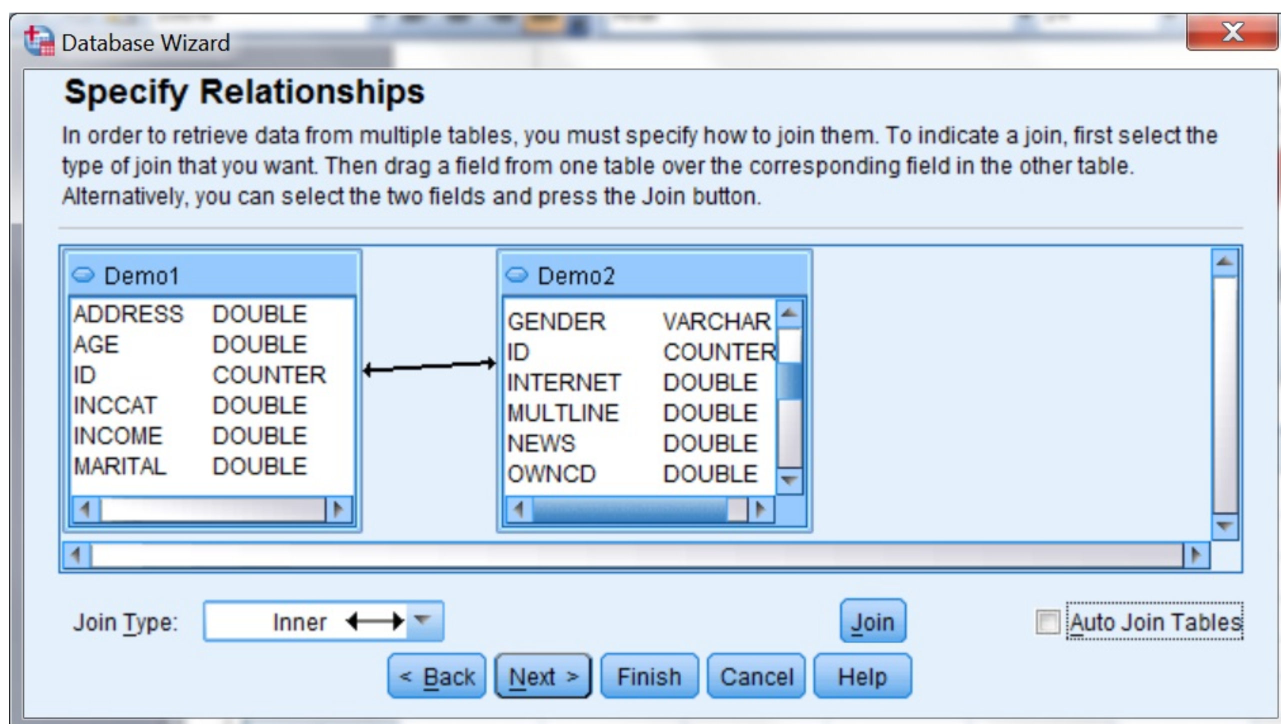
Pentru a pune în evidență modul în care pot fi preluate date din două tabele „Access” între care există o variabilă de legătură am segmentat fişierul „Demo” în două părți, „Demo1” şi „Demo2” păstrând corespondența cazurilor prin intermediul caracteristicii pe care am numit-o „ID” (care este prezentă în ambele fişiere şi permite relaționarea datelor ce se referă la aceleaşi unități statistice). În graficul 3.21 e prezentată selecția varaibilelor, fără modificări în ordinea acestora (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 3.21 Importul variabilelor din „Access”



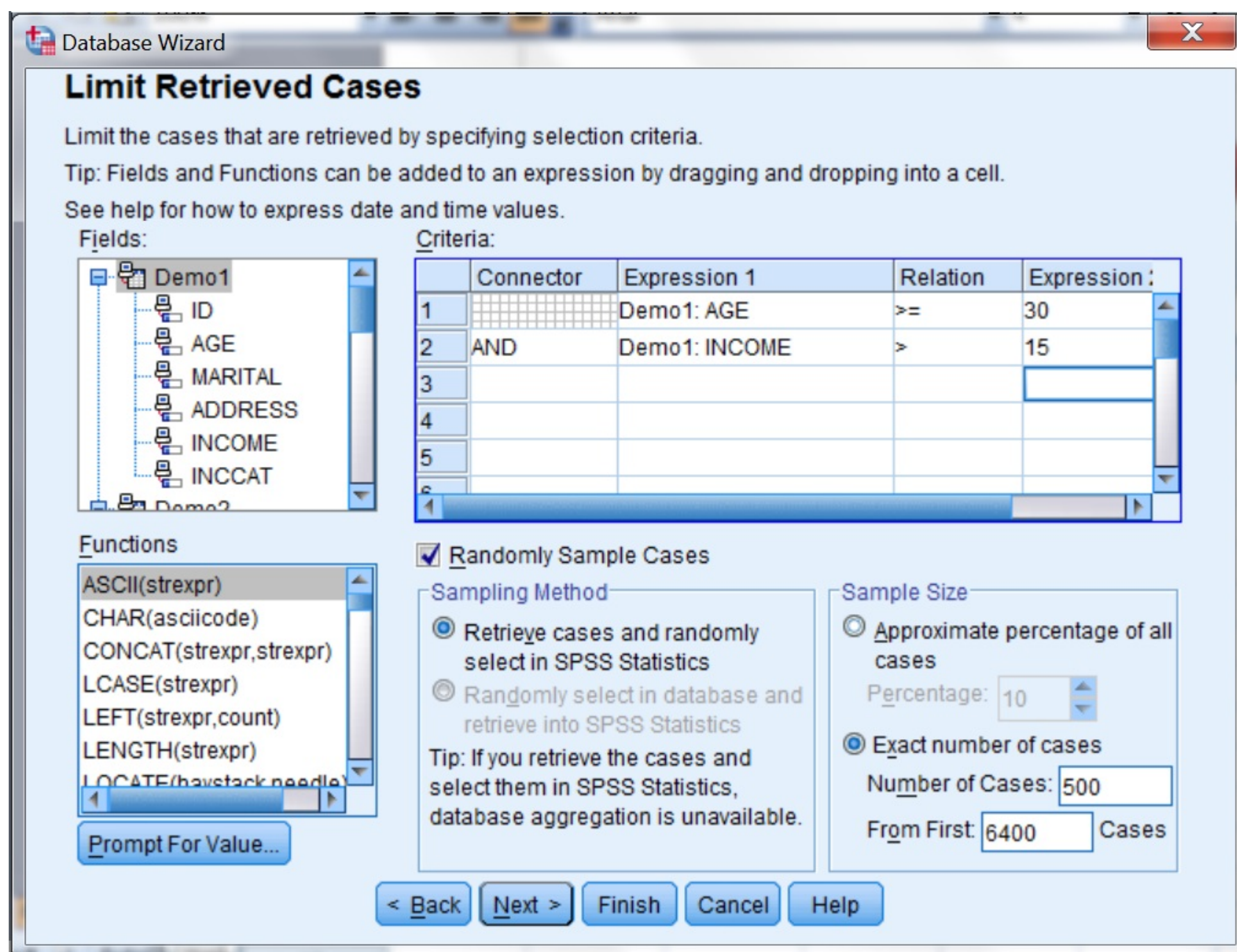
Pentru relaționarea corectă a celor două tabele am apelat la opțiunea „Join” (cunoscută din exemplul foilor de lucru „Excel”); am selectat cele două variabilele cu titulatura „ID” și am ales „Join Type: Inner” (se păstrează doar cazurile unde există valori comune pentru respectivele variabile). Următorul pas îl constituie stabilirea unei (unor) condiții logice pentru selecția cazurilor (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 3.22 Stabilirea legăturii dintre tabelele importate în SPSS



În fereastra de dialog din graficul 3.23 am pus o condiție logică prin care am stabilit că în baza de date SPSS vor fi selectate doar pesoanele care, în același timp, au vârsta de cel puțin 30 de ani (AGE ≤ 30) și posedă un venit anual care este mai mare de 15000 de dolari (INCOME > 15) (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 3.23 Selectarea cazurilor din cele două tabele „Access"



Am apelat și la o selecție aleatoare a 500 de cazuri din totalul de 6400 (identificat prin deschiderea bazei de date „demo" cu programul „Access"). Conform atenționării din fereastră, dacă preluăm cazuri și apoi le selectăm cu ajutorul programului SPSS Statistics, agregarea inițială de date nu va mai fi accesibilă. Evident, dacă dorim să păstrăm și baza de date agregată completă, putem renunța la opțiunea de eșantionare aleatoare. Ulterior, puteam folosi facilitatea existentă în SPSS „Data, Select, Random sample of cases" care permite tratarea datelor identic cu opțiunile existente în partea din dreapta-jos a ferestrei din graficului 3.23. (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

În plus există posiblitatea ca unitățile statistice (cazurile) de interes să fie tratate în mai multe feluri:

– Cazurile neselectate să fie ignorate (filtrate) de programul SPSS în analizele de date realizate după „filtrare"
– Cazurile selectate să fie copiate într-o nouă bază de date
– Cazurile neselectate să fie șterse definitiv

Rezultatul selecției care s-a definit în fereastra „Limit retrived cases” (graficul 3.23) este prezentat sub forma bazei de date SPSS din graficul 3.24. Se poate observa că numărul de cazuri selectate a fost mai mic decât cel indicat de noi. Explicația constă în faptul că valoarea 6400, pe care am indicat-o în fereastra de dialog (graficul 3.23) reprezenta numărul total de cazuri din baza de date; acest număr s-a diminuat prin condiția logică pe care am aplicat-o până la valoarea 5173 (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 3.24 Baza agregata SPSS cu esantion diminuat (405 cazuri)

*Untitled11 [DataSet11] - IBM SPSS Statistics Data Editor

399 : INCOME 42.00 — Visible: 29 of 29 Variables

	ID	AGE	MARITAL	ADDRESS	INCOME	INCCAT	ID1	CARCAT
397	4932	30	.00	7.00	33.00	2.00	4932	2.00
398	6133	30	.00	.00	38.00	2.00	6133	2.00
399	2541	30	.00	8.00	42.00	2.00	2541	2.00
400	5356	30	.00	5.00	43.00	2.00	5356	2.00
401	485	30	1.00	4.00	47.00	2.00	485	2.00
402	1322	30	.00	3.00	58.00	3.00	1322	2.00
403	5086	30	1.00	5.00	61.00	3.00	5086	3.00
404	5667	30	.00	.00	81.00	4.00	5667	3.00
405	5123	30	.00	.00	83.00	4.00	5123	3.00
406								

Graficul 3.25 Baza agregata SPSS cu esantion complet (498 cazuri)

*Untitled14 [DataSet14] - IBM SPSS Statistics Data Editor

498 : AGE 30.00 — Visible: 29 of 29 Variables

	ID	AGE	MARITAL	ADDRESS	INCOME	INCCAT	ID1	CARCAT
490	2979	30.00	.00	3.00	47.00	2.00	2979	2.00
491	3299	30.00	.00	5.00	41.00	2.00	3299	2.00
492	3424	30.00	.00	1.00	30.00	2.00	3424	2.00
493	4239	30.00	.00	6.00	31.00	2.00	4239	2.00
494	4531	30.00	.00	11.00	39.00	2.00	4531	2.00
495	4877	30.00	1.00	1.00	45.00	2.00	4877	2.00
496	5321	30.00	.00	4.00	34.00	2.00	5321	2.00
497	5904	30.00	1.00	7.00	31.00	2.00	5904	2.00
498	6060	30.00	.00	9.00	22.00	1.00	6060	1.00
499								

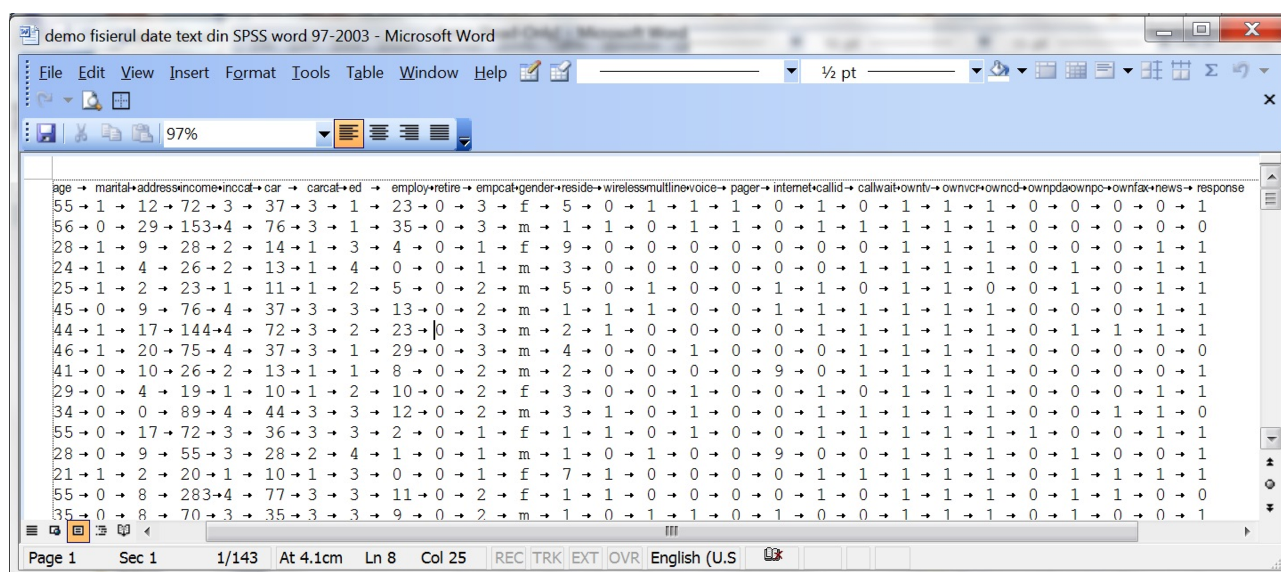
În baza de date din graficul 3.24 eșantionul extras are doar 405 cazuri datorită volumului disponibil selecției inferior celui declarat pe când în graficul 3.25 vedem că mărimea eșantionului e corectă, cu excepția, probabil, a unor date lipsă care au afectat aplicarea condiției logice de selectare a cazurilor (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

3.4 Citirea din SPSS, a datelor din fișierele text

Fișierele text reprezintă o sursă importantă de date având în vedere faptul că multe programe de calcul tabelar sau de gestionare a bazelor de date pot să salveze informațiile sub diferite forme de fișiere text (The Basics: SPSS for Windows, 2002).

Pentru a putea importa în SPSS baze de date sub format text este foarte important să cunoaștem modul în care sunt organizate respectivele date. Vom apela la exemplul de fișier text (cu structură de bază de date) existent în „Program Files”, „IBM”, „SPSS”, „Statistics”, „20”, „Samples”, „English”, denumit „Demo” și având extensia .txt. Este esențial să cunoaștem modul în care este structurată informația la nivelul acestui fișier și, ca atare, este preferabil ca inițial să fie deschis cu ajutorul programului „Microsoft Word”, selectându-se opțiunile de vizualizare a tuturor caracterelor ascunse, în special a spațiilor existente între acestea (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Graficul 3.26 Structura datelor din fisierul text „Demo”



age	marital	address	income	inccat	car	carcat	ed	employ	retire	empcat	gender	reside	wireless	multline	voice	pager	internet	callid	callwait	owntv	ownvcr	owncd	ownpda	ownpc	ownfax	news	response
55	1	12	72	3	37	3	1	23	0	3	f	5	0	1	1	1	0	1	0	1	1	1	0	0	0	0	1
56	0	29	153	4	76	3	1	35	0	3	m	1	1	0	1	1	0	1	1	1	1	1	0	0	0	0	0
28	1	9	28	2	14	1	3	4	0	1	f	9	0	0	0	0	0	0	0	1	1	1	0	0	0	1	1
24	1	4	26	2	13	1	4	0	0	1	m	3	0	0	0	0	0	0	1	1	1	1	0	1	0	1	1
25	1	2	23	1	11	1	2	5	0	2	m	5	0	1	0	0	1	1	0	1	1	0	0	1	0	1	1
45	0	9	76	4	37	3	3	13	0	2	m	1	1	1	0	0	1	1	1	1	1	1	0	0	0	1	1
44	1	17	144	4	72	3	2	23	0	3	m	2	1	0	0	0	0	1	1	1	1	1	0	1	1	1	1
46	1	20	75	4	37	3	1	29	0	3	m	4	0	0	1	0	0	0	1	1	1	1	0	0	0	0	0
41	0	10	26	2	13	1	1	8	0	2	m	2	0	0	0	0	9	0	1	1	1	1	0	0	0	0	1
29	0	4	19	1	10	1	2	10	0	2	f	3	0	0	1	0	0	1	0	1	1	1	0	0	0	1	1
34	0	0	89	4	44	3	3	12	0	2	m	3	1	0	1	0	0	1	1	1	1	1	0	0	1	1	0
55	0	17	72	3	36	3	3	2	0	1	f	1	1	0	1	0	0	1	1	1	1	1	1	0	0	1	1
28	0	9	55	3	28	2	4	1	0	1	m	1	0	1	0	0	9	0	0	1	1	1	0	1	0	0	1
21	1	2	20	1	10	1	3	0	0	1	f	7	1	0	0	0	0	1	1	1	1	1	0	1	1	1	1
55	0	8	283	4	77	3	3	11	0	2	f	1	1	0	0	0	0	1	0	1	1	1	0	1	1	0	0
35	0	8	70	3	35	3	3	9	0	2	m	1	0	1	1	0	1	0	0	1	1	1	0	1	0	0	1

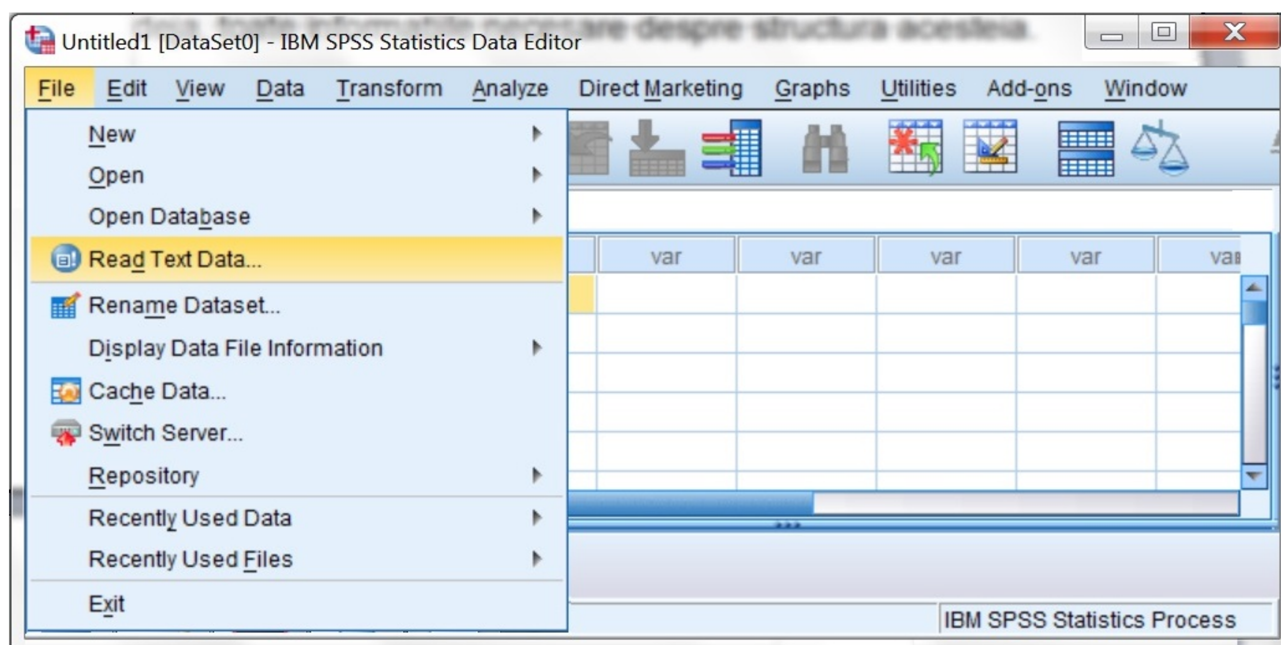
Folosind programul „Microsoft word” am deteminat faptul că datele apar pe coloane și am stabilit că în fiecare coloană există un nume al variabilei. De asemenea, am putut detetermina faptul că datele sunt separate unele de altele prin intermediul tab-urilor (am identificat 28 de variabile).

Vom prezenta, în continuare, pașii din SPSS necesari pentru importarea bazei de date format text („demo”) în condițiile în care deținem, deja, toate informațiile necesare despre structura acesteia (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

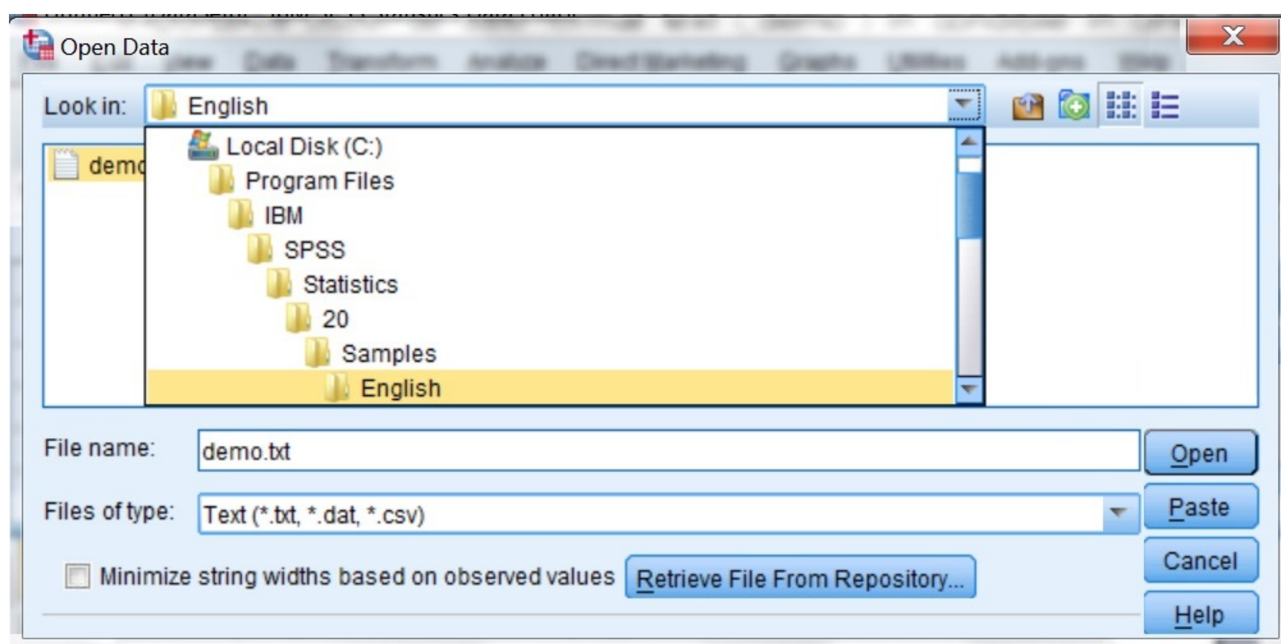
Primul pas necesar în demersul nostru îl constituie selectarea comenzii „File, Read Text Data” care ne va conduce la o fereastră de dialog prin intermediul căreia vom alege calea prin care se ajunge la fișierul text vizat.

În cazul nostru, după cum am arătat deja, calea de acces spre fișierul text este „Program Files”, „IBM”, „SPSS”, „Statistics”, „20”, „Samples”, „English”. În căsuța „Files of type” se va alege Text (cu extensia „.txt”, „.dat” sau „.csv”) iar în „File name” se va selecta „demo.txt”. În plus, se poate opta pentru minimizarea coloanelor cu caractere string (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Graficul 3.27 Comanda de inițiere a importării bazei de date de tip text



Graficul 3.28 Indicarea căii către fișierul text (structurat ca bază de date)



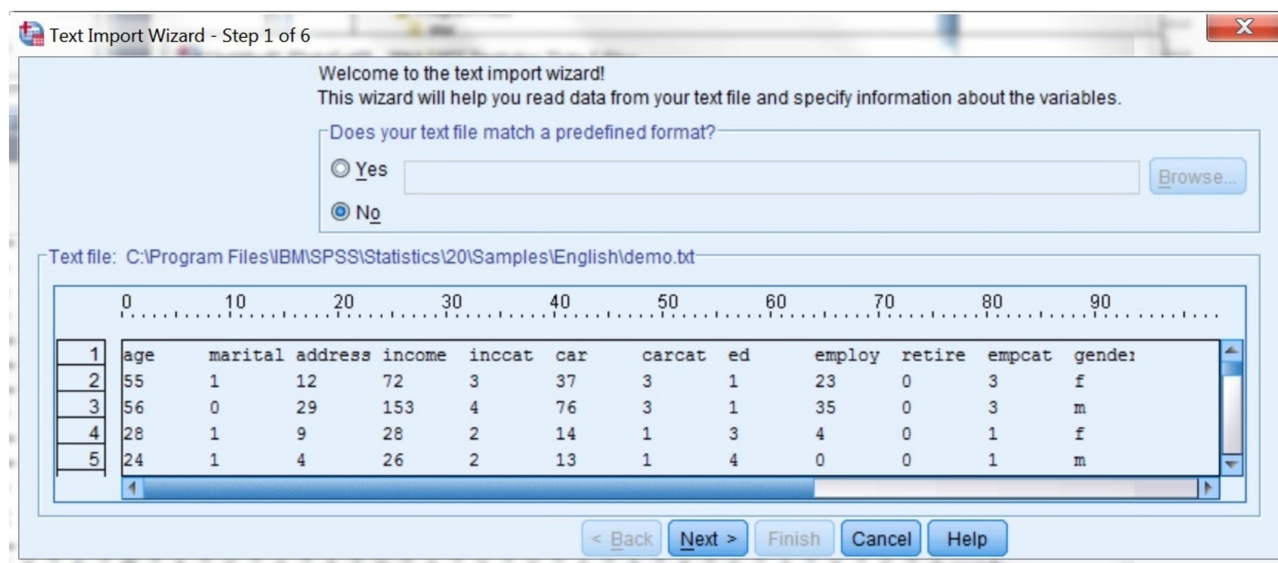
După deschiderea fișierului text „demo” (cu structura de bază de date) va apărea o nouă fereastră (graficul 3.29) în care programul SPSS va oferi informații despre modul îm care interpretează noile informații (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

De notat faptul că preluarea datelor s-ar putea realiza și activând un fișier „TextWizard Predefined Format”, având extensia .tpf, care putea fi creat în urma salvării unei proceduri de preluare a informațiilor dintr-un fișier având aceeași structură ca și cel vizat în momentul de față.

Dacă nu dispunem de o procedură de preluare a datelor care să fie aplicabilă fisierului „demo”, se va alege varianta de răspuns „No” la întrebarea „Does your text file match a predefined format?” (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

În graficul 3.29 apare o imagine a variabilelor identificate de SPSS sub forma unor coloane de date care au și o ipotetică denumire (primul rând al tabelului). Putem observa, în plus, că separarea datelor și denumirea variabilelor prezentată în graficul 3.29 corespund structurii datelor pe care am identificat-o în graficul 3.26 (totuși, unele denumiri sunt decalate).

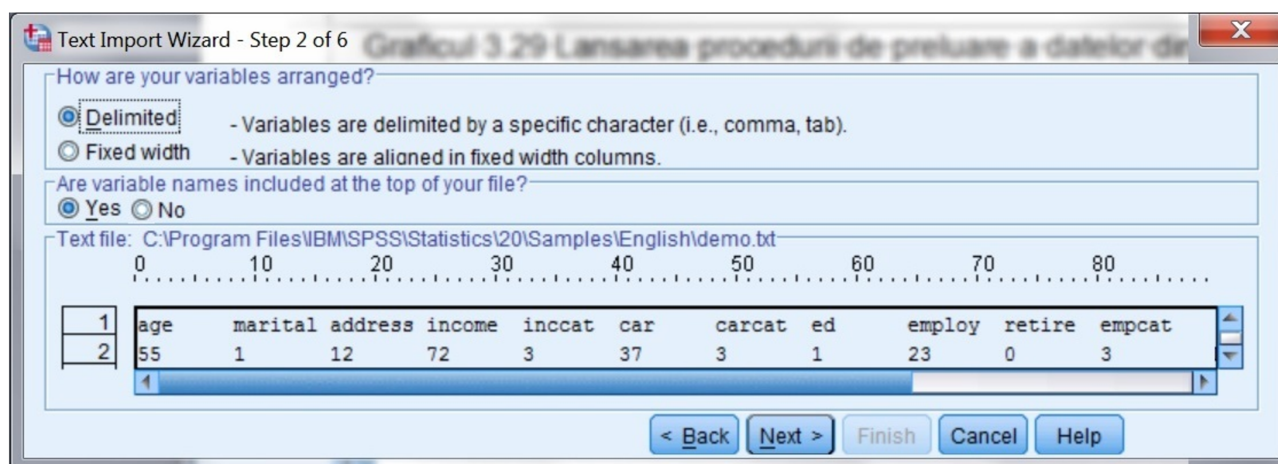
Graficul 3.29 Lansarea procedurii de preluarea datelor dintr-un fișier text



Pașii următori ai procedurii presupun solicitarea, din SPSS, a unor informații despre modul în care sunt organizate variabilele în fisierul de tip text pentru a se asigura transferul corect și complet al datelor.

În pasul al doilea necesar transferării datelor din fișierul text trebuie stabilit dacă variabilele sunt delimitate între ele print caractere specifice, cum ar fi tabulatori, virgule, spații sau alte caractere care pot fi folosite pentru a separa valorile variabilelor. Alternativa ar fi ca datele să fie delimitate unele de altele implicit, fără utilizarea caracterelor specifice ca separator. În acest caz, varianta care trebuie aleasă e „Fixed width" Valorile fiecărei variabile sunt înregistrate în coloanele specific destinate (fapt valabil pentru fiecare caz / linie din fișierul de date) fără necesitatea delimitării respectivelor variabile (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

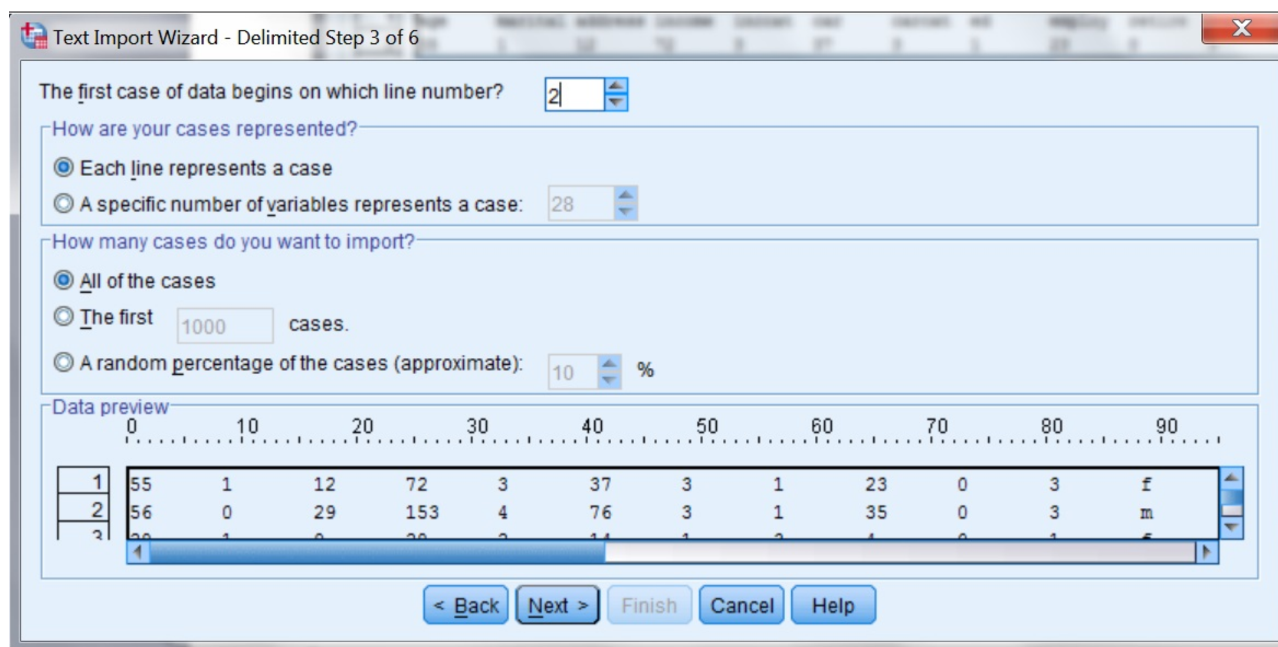
Graficul 3.30 Al doilea pas pentru preluarea datelor dintr-un fișier text



Următoarea informație care trebuie transmisă vizează existența numelor pentru variabile, poziționate în partea de sus a coloanei asociată fiecăreia dintre ele. În cazul nostru, după cum se poate vedea în graficul 3.26, în partea de sus a fiecărei coloane se află denumirea acesteia.

Pasul al treilea al procedurii de importare a datelor de tip text în SPSS presupune specificarea rangului liniei tabelului în care apar datele primului caz din baza de date (prima linie, a doua linie sau a treia, etc.). În cazul particular pe care îl analizăm, datele din primul caz al bazei de date apar în linia a II-a, prima linie fiind destinată numelui variabilei (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 3.31 Al treilea pas pentru preluarea datelor dintr-un fișier text



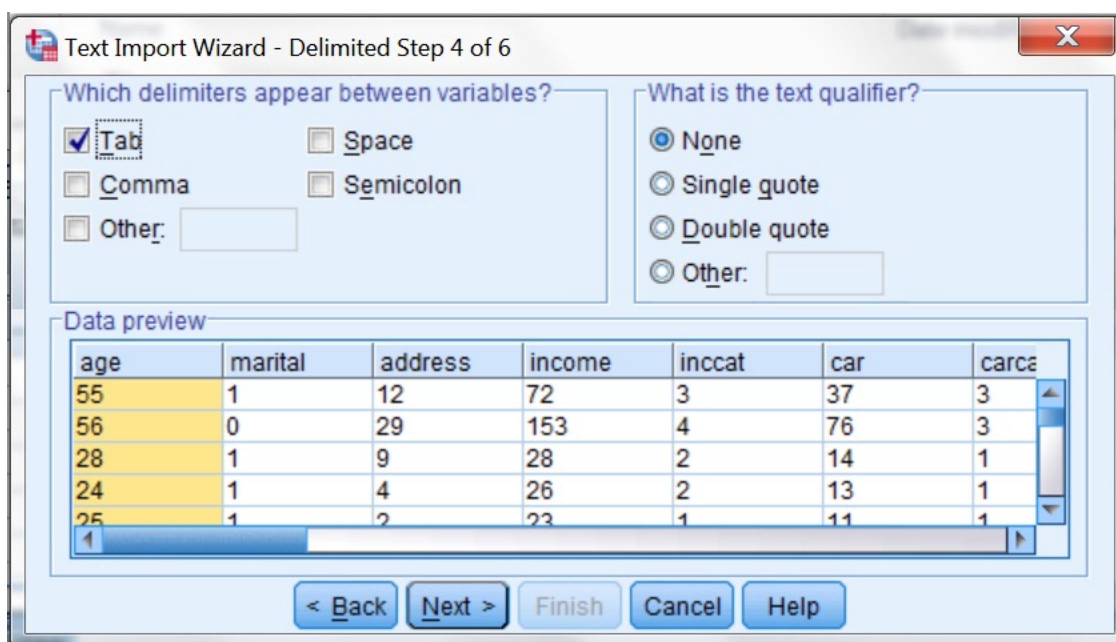
O altă precizare necesară vizează modul în care sunt reprezentate cazurile. În exemplul nostru, fiecare linie reprezintă un caz la nivelul bazei de date. Alternativa oferită de SPSS este ca un anumit număr de variabile să reprezinte un caz. Aceasta ar însemna că în căsuța „A specific number of variables represent a case” să stabilim ca programl SPSS să citească un anumit număr „X” de elemente ale primei linii ca fiind cazul 1, urmatoarele X” elemente ca fiind cazul 2, etc. Citirea se face succesiv pe linii și nu are importanță dacă un caz începe la sfârșitul unei linii și se continuă pe cea următoare.

Condiția esențială, în această procedură, este ca fiecărui caz să-i fie alocate exact numărul de valori specificate (nonrăspunsurile, dacă apar, trebuie marcate cu anumite caractere specifice). În caz contrar, citirea fișierul text se va face eronat (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

În fereastra din graficul 3.31 există posibiliitatea de a alege numărul de cazuri care se dorește a fi importat, cu 3 opțiuni: toate cazurile, primele......... cazuri respectiv o selecție aleatoare (eșantion) care să reprezinte un procent aproximativ din totalul cazurilor. În exemplul nostru concret am optat pentru selectarea tuturor cazurilor. Celelate două variante sunt utile dacă avem informații despre fișierul text și ne interesează doar primele „X” cazuri sau dacă sunt foarte multe cazuri și dorim să selectăm doar un eșantion aleator (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Dacă vom da click pe butonul „next” al ferestrei din graficul 3.31 vom lansa cel de al patrulea pas al ghidului de importare a datelor tip text („Text Import Wizard”). Prima informație cerută vizează natura caracterelor care delimitează variabilele unele de altele. Opțiunile predeterminate sunt „Tab” (adică tabulatorul, utilizat în exemplului concert pe care îl prezentăm), spațiul (space), virgula („Comma”) respectiv punctul și virgula (Semicolon). Opțiunea „Other” permite indicarea oricărei alte combinații de caractere care pot să delimiteze variabilele. Este important de reținut faptul că lipsa oricăror valori dintre două seturi de caractere definite ca delimitatori ai coloanelor va fi interpretată ca dată lipsă (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Graficul 3.32 Al patrulea pas pentru preluarea datelor dintr-un fișier text

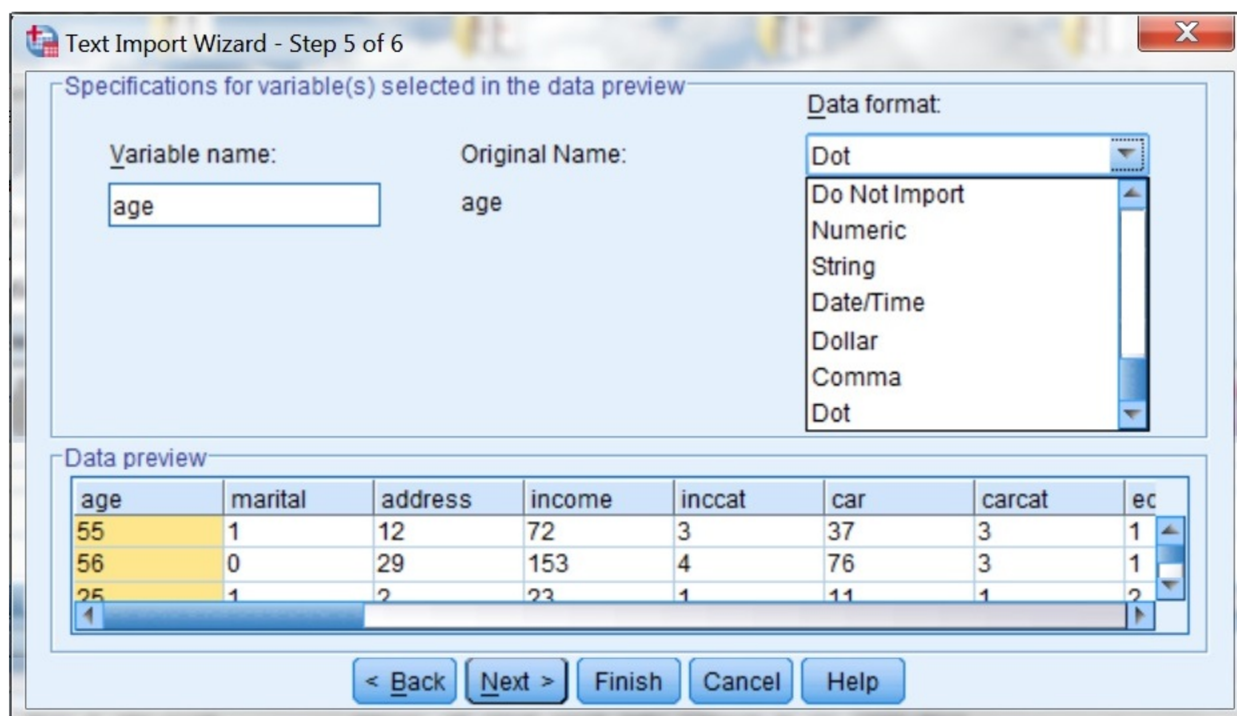


Următoarea informație solicitată de „Text Import Wizard” vizează caracterul care are rolul de a marca începutul și sfârșitul fiecărei valori din fișierul text. Se pot folosi ghilimele simple sau ghilimele duble cu scopul de a nu se face confuzii între carcterele utilizate pentru a scrie / exprima valorile variabilelor și caracterul destinat separării coloanelor unele de celelalte (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

De exemplu, putem folosi virgula ca element de separare între valorile unui tabel text tip bază de date. S-ar putea, însă, ca printre valorile din tabel să avem și o variabilă numerică cu două zecimale pentru care delimitare dintre partea întreagă și cea zecimală să se realizeze tot prin virgulă. În aceste condiții, citirea corectă a datelor se poate realiza doar dacă în prealabii s-a apelat la un separator al valorilor din baza de date de tip text. De exemplu, se pot folosi ghilimele simple sau duble, prezente la începutul și sfârșitul fieărei valori (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

În fereastra din graficul 3.32 se poate opta pentru a defini separatorul valorilor prin ghilimele simple („Simple quote”), duble („Double quote”) sau pentru a indica un alt tip de caracter („Other:”). În exemplul nostru nu există un separator al valorilor variabilelor deoarece delimitarea coloanelor se face prin tabulatori (care nu sunt conținuți în interiorul valorilor din fișierul text și, ca atare, nu riscă să creeze confuzii în raport cu valorile).

Graficul 3.33 Al ciniclea pas pentru preluare a datelor dintr-un fișier text



În fereastra din gaficul 3.33 sunt prezentate opțiunile de formatare pentru citirea variabilelor cu ajutorul mediatorului / expertului de text („Text Wizard"). Acestea sunt:

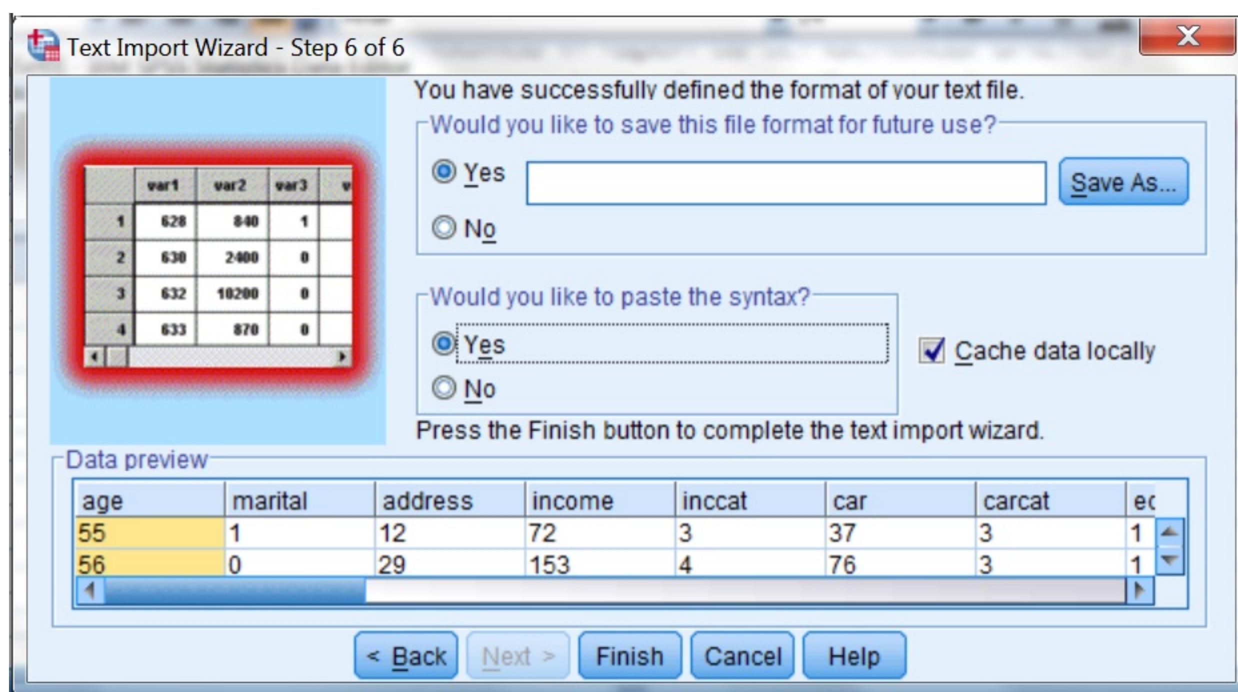
1. **Nu importați**. Se omit unele variabile din fișierul de date importat (acele variabile ce nu se doresc a fi importate vor fi selectate iar în căsuța „Data format" se va alege varianta „Do Not Import")
2. **Numeric**. Valorile valide includ numerele, un semn de plus sau minus și un indicator zecimal.
3. **Şir de caractere** (String). Valorile valide includ orice tip de caractere sau spații libere. Pentru fișierele unde valorile sunt delimitate prin separatori se poate specfica numărul de caractere destinat valorilor alfanumerice până la un maxim de 32.767. Implicit, programul SPSS va stabili numărul de maxim de caractere ca fiind egal cu al celei mai lungi valori „string" întâlnită pentru variabila selectată în primele 250 de rânduri ale fișierului. Pentru fișierele unde lățimea e fixă, numărul de elemente al șirului de caractere este definit prin poziția caracterelor care separă variabilele (a se vedea pasul 4).
4. **Data / ora**. Valorile valide includ datele cu formatul general zz-ll-aaaa, ll-zz-aaaa, zz.ll.aaaa, ll.zz.aaaa, zz/ll/aaaa, ll/zz/aaaa precum și o varietate de alte formate de dată, unele combinate cu ora, minutul și secunda.
5. **Dolar**. Valorile valide sunt numerele cu semnul opțional **„$"** și cu virgulele opționale ca separatoare pentru ordinele de mărime (mii, milioane, etc.).
6. **Virgulă** (Coma). Valorile valide includ numere în care punctul este indicator zecimal și virgula este separator pentru ordinele de mărime (mii, milioane).
7. **Punct** (Dot). Valorile valide includ numere care folosesc virgula ca indicator zecimal și punctul ca separator pentru diferitele ordine de mărime (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Este important să ținem seama de următorul aspect: valorile care conțin caractere nevalide pentru formatul declarat vor fi considerate date lipsă. Valorile care conțin oricare dintre caracterele declarate ca delimitatori ai valorilor variabilelor vor fi tratate ca valori multiple.

În exemplul pe care îl analizăm, au fost identificate 28 de variabile, dintre care 27 de tip „Numeric" și una de tip String („Gender", având numărul de caractere egal cu 1). Notăm faptul că pentru fiecare varaibilă în parte trebuie precizat numele (dacă se modifică în raport cu cel identificat anterior) și trebuie indicat formatul datelor. E important să se verifice dacă varianta implicită propusă de SPSS corespunde cu structura datelor din fișierul text (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

În finalul demersului nostru, vom obține o fereastră care permite salvarea fișierului de tipul „TextWizard Predefined Format", având extensia .tpf și care va putea fi utilizat ulterior pentru transferul în SPSS a altor fișiere text, având structură identică cu cel vizat de noi în actualul exemplu.

Graficul 3.34 Salvarea fisierului „TextWizard" și a sintaxei utilizate



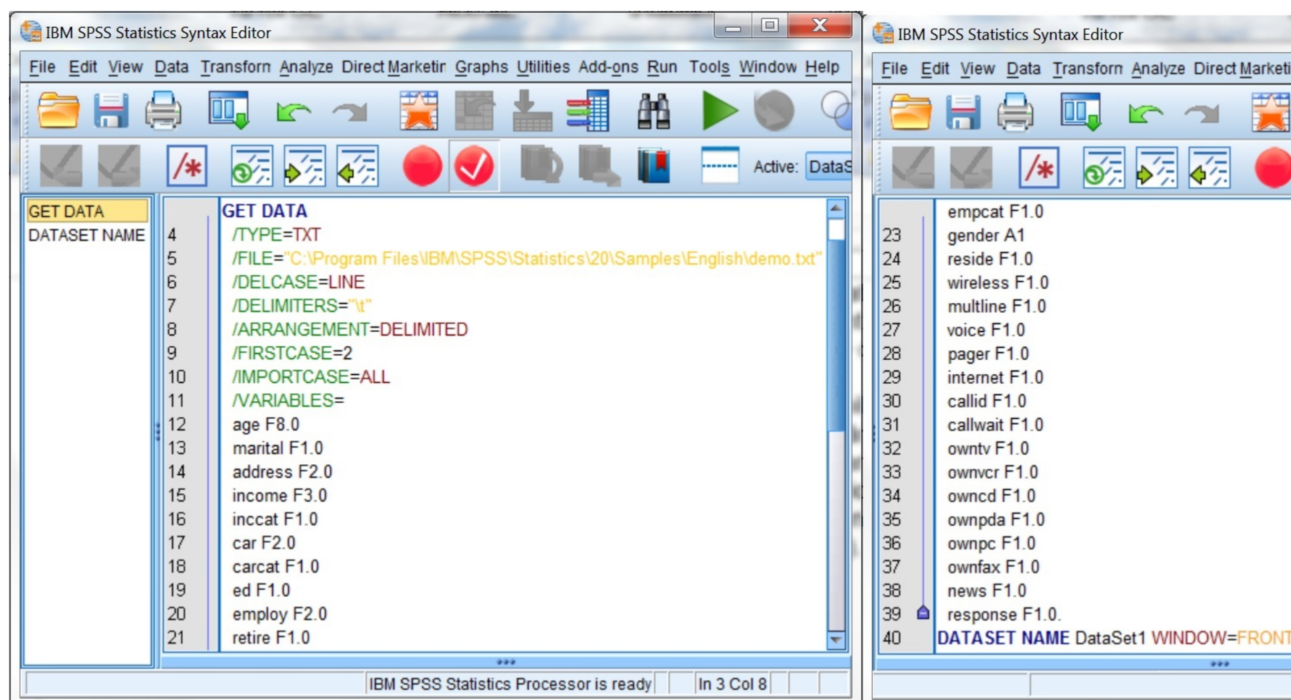
Aceeași fereastră oferă posibilitatea transferului succesiunii comenzilor SPSS formulate în timpul pregătirii transferului datelor din fișierul text (având structură tabelară).

În graficul 3.35 este prezentat fișierul tip sintaxă în care apar comenzile SPSS, începând cu „GET DATA", specificarea tipului de fișier „TXT" și indicarea succesiunii de foldere prin care se poate accesa fișierul text organizat ca bază de date („Program Files \ IBM \ SPSS \ 20 \ English \ demo.txt)

În continuare se specifică faptul că fiecare linie reprezintă un caz și apoi că delimitatorii valorilor (coloanelor) succesive sunt tabulatorii. Se indică, în continuare, că modul de aranjare a datelor în tabelul text este de tip delimitat, că primul caz începe pe a doua linie a tabelului (prima fiind destinată descrierii variabielelor) și că se vor importa toate variabilele (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Fișierul sintaxă conține, în continuare, o descriere a numelui variabilei și a formatului atribuit acesteia în viitoarea bază de date SPSS. „F” arată că variabila e de tip numeric, cifra următoare definește maximul de caractere alocat iar valoarea de după punct exprimă numărul zecimalelor. Pentru variabila alfanumerică „gender” în fișierul sintaxă apare simbolul „A1” ceea ce desemnează o caracteristică de tipul „string” având alocat un singur caracter pentru fiecare caz.

Graficul 3.35 comenzile SPSS de preluare a datelor tip text (fișierul sintaxă)



Graficul 3.36 Baza de date SPSS creată din fișierul text „demo” (6400 cazuri)

*Untitled3 [DataSet2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 28 of 28 Variables

	age	marital	address	income	inccat	car	carcat	ed	employ	retire	empcat	gender	reside	wireless	multline
6394	62	1	25	122	4	59	3	2	30	0	3	m	2	0	1
6395	52	0	12	44	2	22	2	3	7	0	2	f	1	1	1
6396	56	0	22	207	4	85	3	1	39	0	3	m	1	0	0
6397	40	0	21	60	3	31	3	4	6	0	2	f	2	0	0
6398	34	0	3	59	3	29	2	4	9	0	2	m	6	1	0
6399	26	1	3	44	2	22	2	4	2	0	1	m	3	0	0
6400	38	0	4	62	3	31	3	2	13	0	2	m	2	0	0
6401															

IBM SPSS Statistics Processor is ready

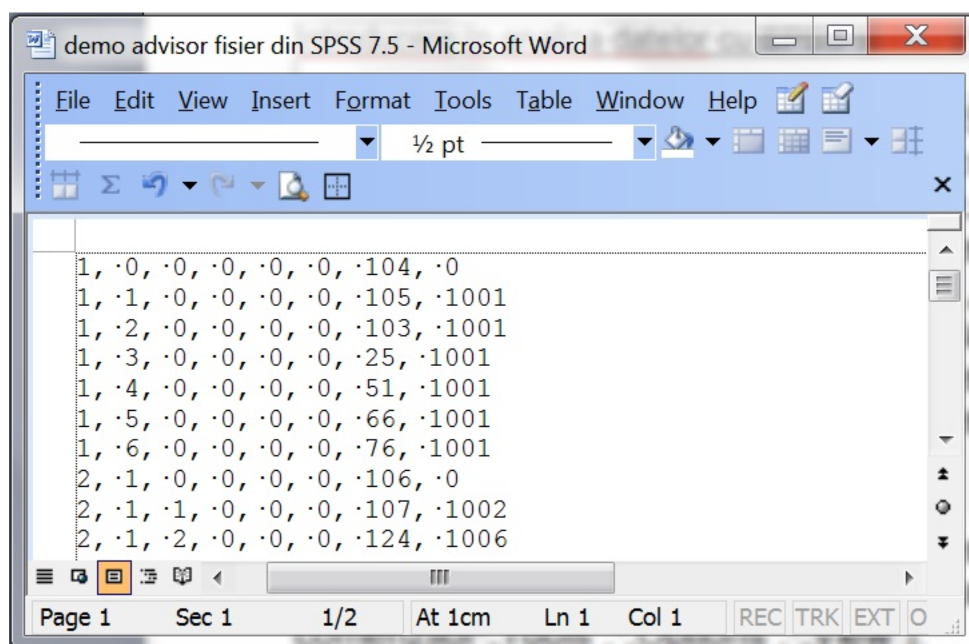
În graficul 5.36 s-a obținut forma finală a bazei de date SPSS în care s-au transferat cele 6400 de cazuri din fișierul text, subsumând 28 de variabile (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Notăm faptul că instrucțiunile SPSS de transfer a datelor din fișierul text în baza de date proprie respectivului soft statistic apar și în mod implicit, în fișierul de tip „output” (lansat automat de SPSS). Ele pot fi preluate și de aici, prin „copy” simplu și „paste” într-un fișier de tip sintaxă. Dacă în respectivul fișier de tip sintaxă vom da comanda „Run”, întegul proces de preluare a datelor din fișierul text va fi reluat, până la obținerea bazei de date SPSS (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Vom mai da un exemplu de transfer al unui fișier de tip text, de această dată având extensia „.dat.” Vom prezenta succint noul demers, având în vedere similitudinile ci cel precedent.

Pasul cel mai imporant este acela de identificare a structurii datelor din fișierul pe care îl vizăm prin deschiderea acestuia cu programul Microsoft Word (selectând, după cum am mai spus, vizualizarea caracterelor ascunse din interiorul textului fișierului cu ajutorul comenzilor „Tools”, „Options”, „View”).

Graficul 3.37 Structura datelor din fisierul „Advisor”

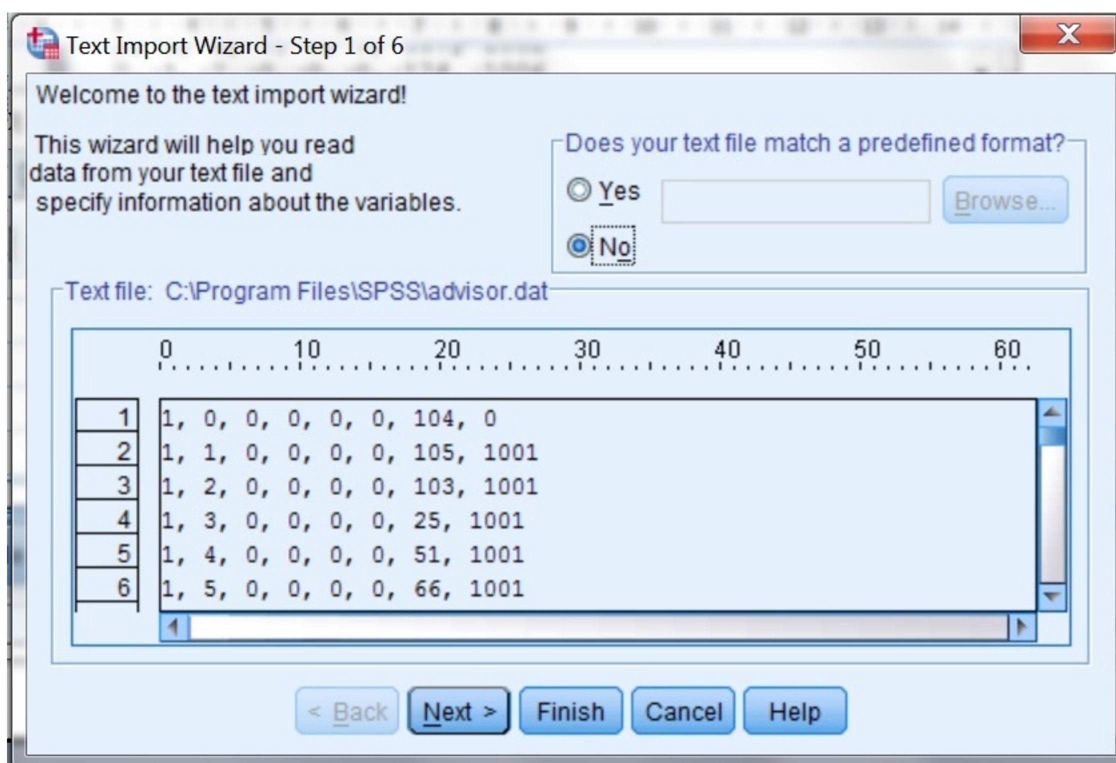


În fisierul 3.37 se poate observa că datele sunt organizate pe coloane (fiecare caz reprezintă o linie a tabelului) și că între coloane exist separatori. De fapt, oricare două valori succesive ale unui caz sunt separate de o virgulă urmată de un spațiu liber. Respectivul spațiu liber este reprezentat în graficul 3.37 printr-un punct deoarece s-a optat pentru vizualizarea caracterelor ascunse (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

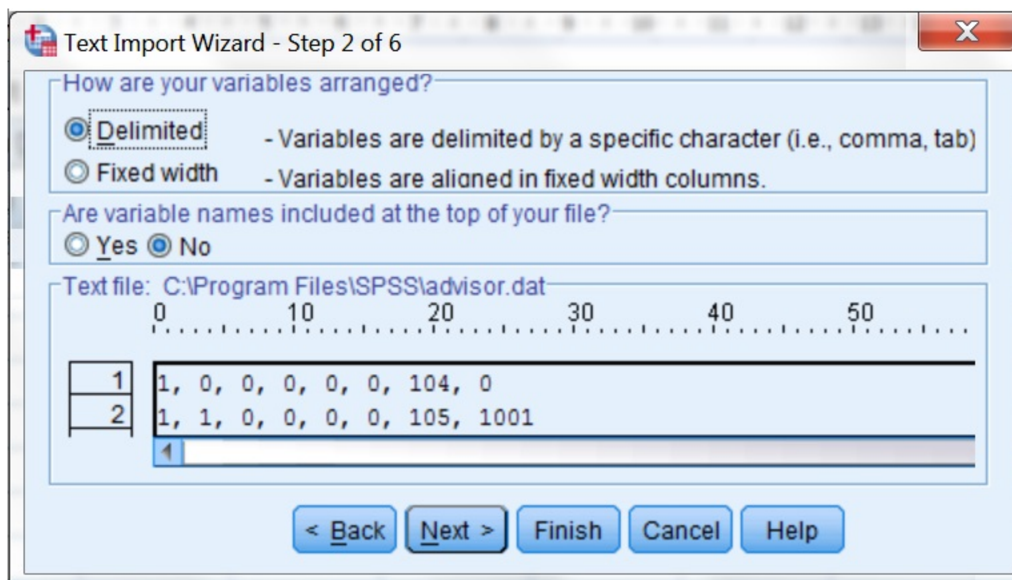
În graficul 3.38 este prezentat primul pas al importării fișierul text sub supervizarea „expertului” prezent în programul SPSS („Text import wizard”). În căsuța care se referă la existența unui format predeterminat pentru importarea fișierului text vom răspunde negativ, ceea ce înseamnă că nu am mai importat fișiere având structuri identice (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Al doilea pas al importării fișierului text (conținând date stastistice) este prezentat în graficul 3.39. Aici vom arăta faptul că variabilele (ale căror date apar pe coloane) sunt delimitate de un caracter specific și, în plus, vom indica faptul că numele variabilelor nu sunt specificate în partea de sus a coloanelor.

Graficul 3.38 Pasul 1; transfer în SPSS al fisierului text (extensia .dat)



Graficul 3.39 Pasul 2; transferul în SPSS a fisierului text (.dat)



Următorii 2 pași ai transferului de date dinspre fisierul text spre SPSS sunt prezentați în graficul 3.40. În primu rând, vom specifica faptul că primul caz conține date statistice începând cu prima linie (în dreptul întrebării „The first case of data begins on which line number?” notăm cifra 1). Apoi arătăm că fiecare linie reprezintă un caz al bazei de date („Each line represents a case”).

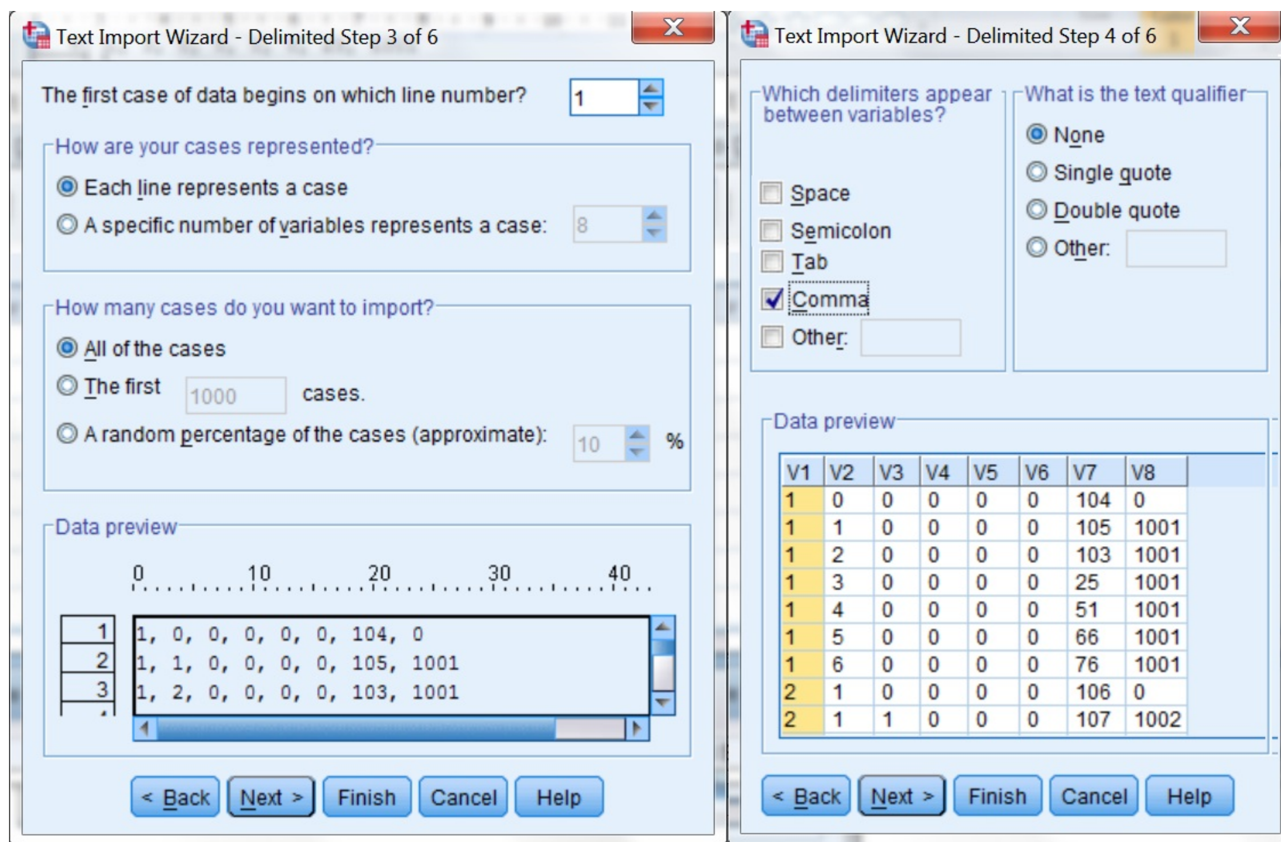
În căsuța de mai jos a ferestrei (graficul 3.40) se poate opta pentru importarea tuturor valorilor din fișierul text, pentru indicarea numărului primelor cazuri care vor fi selectate sau a ponderii, din totalul cazurilor, pe care o reprezintă volumul unui eșantion care va fi extras aleator (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

În pasul 4 al procesului de importare a datelor (partea dreaptă a graficului 3.40) se indică acel caracter care delimitează două variabile succesive. În cazul nostru, situația este atipică deoarece apar, împreună, o virgulă („comma”) urmată de un spațiu („Space”).

Dacă alegem ambele simboluri, SPSS va dubla numărul de coloane (fiecărei coloane cu date îi va corespunde una vidă).

Dacă, însă, delimitatorul variabilelor va fi definit ca virgulă, fiecărei coloane de date îi va corespunde o singură variabilă în baza de date SPSS (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 3.40 Pasul 3 și 4; transferul în SPSS a fisierului text (.dat)

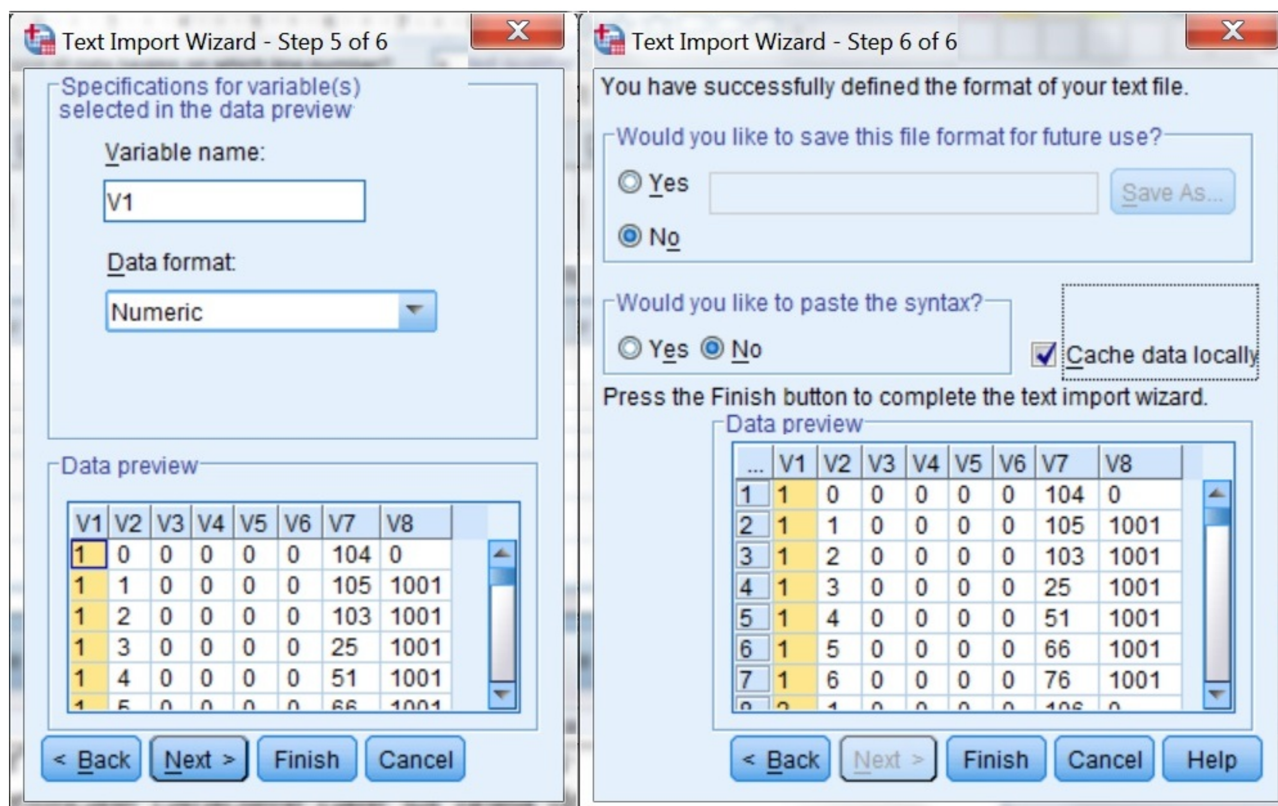


Pentru simbolul utilizat în delimitarea valorilor din fișierul text („text qualifier”) nu s-a specificat nimic având în vedere că în graficul 3.37 nu au fost identificate caractere care să joace rolul respectiv (textul conținea doar valori numerice fără zecimale, în care nu s-a folosit virgula pentru a delimita ordinele de mărime (cifra miilor, în speță).

Ultimele două ferestre care apar în demersul de citire a unui fișier text din programul SPSS sunt prezentate în graficul 3.41. Pasul 5 presupune ca fiecare variabilă să fie definită în funcție de specificul acesteia: numerică, șir de caractere, numerică cu punct zecimal și virgulă pentru delimitarea miilor, etc. În cazul nostru, toate variabilele sunt numerice întregi și nu folosesc virgula sau alt simbol pentru a delimita miile respectiv ordinele de mărime (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

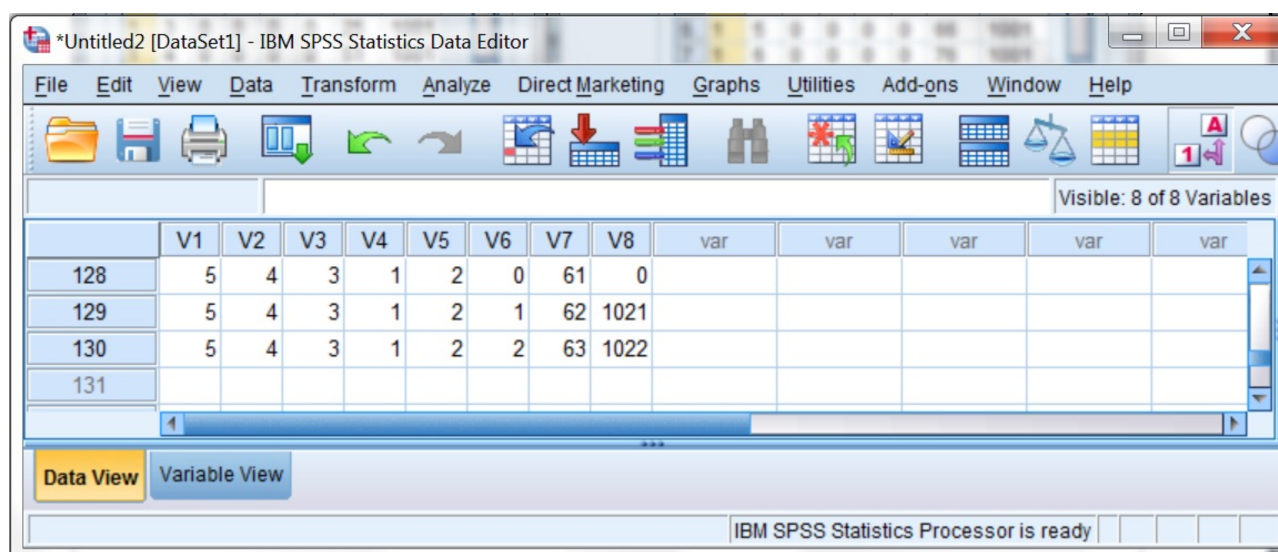
În finalul demersului de importare, în SPSS, a bazei de date de tip text (cu extensia .dat) putem salva fișierul „TextWizard Predefined Format”, având extensia .tpf, pentru a-l folosi în transferul structurilor de date identice (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 3.41 Pasul 5 şi 6; transferul în SPSS a fisierului text (.dat)



Ca şi în cazul prezentat precedent, putem salva întrun fişier sintaxă comenzile care au fost date în SPSS şi care ne permit să reluăn toţi paşii care au fost făcuţi pentru preluarea datelor din fişierul text printr-o singură comandă „Run,” care lansează succesiunea liniilor de program SPSS (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 3.42 Baza de date SPSS obţinută (130 de cazuri şi 8 variabile)



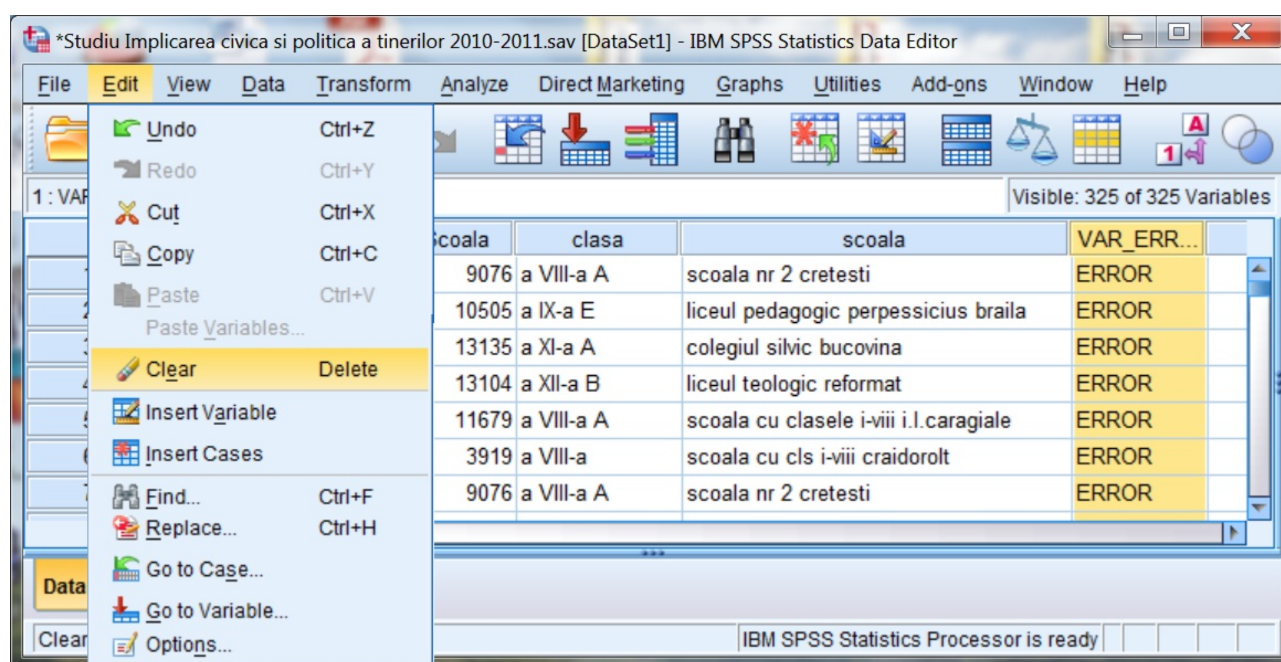
În continuare vom părăsi problematica transferului de date în SPSS, deși nu am epuizat-o, pentru a ne concentra asupra facilităților oferite de fereastra „Data Editor”, principala sursă a gestionării a datelor statistice.

4. MENIUL EDIT ÎN SPSS ŞI PSPP

4.1 Facilitățile EDIT, Cut, Copy, Paste, Clear și Insert în SPSS și PSPP

Similar aplicațiilor windows, Meniul **EDIT**, atât în SPSS cât și în PSPP, oferă o serie de facilități legate de operațiile cu fișiere în sensul modificării conținuturilor acestora (sau copierii lor, prin comanda **EDIT COPY**). Opțiunea **EDIT CLEAR** e utilizată atunci când se dorește înlăturarea unor părți din fișier care nu mai sunt utile. Prin „Edit Clear" se pot şterge variabile din baza de date în întregime sau doar conținutul unui anumit număr de celule ale acestora. În graficul 4.1 este pus în evidență modul în care se poate șterge variabila cu numele VAR_ERROR prin selectarea ei urmată de comanda „Clear" (A Handbook of Statistical Analyses using SPSS).

Graficul 4.1 Ştergerea unei variabile din baza de date cu CLEAR în SPSS



Instrucțiunea **EDIT CUT** e folosită, pentru a extrage informația dintr-o anumită zonă și a-i atribui o nouă destinație (în același fișier sau în altul). Folosind comanda „Cut" vom şterge informația din zona selectată inițial și o vom muta-o acolo unde dorim. Pasul următor constă în inițierea comanzii Paste (procedura va avea succes dacă structura selectată în care se va încerca transferul va fi compatibilă cu cea inițială).

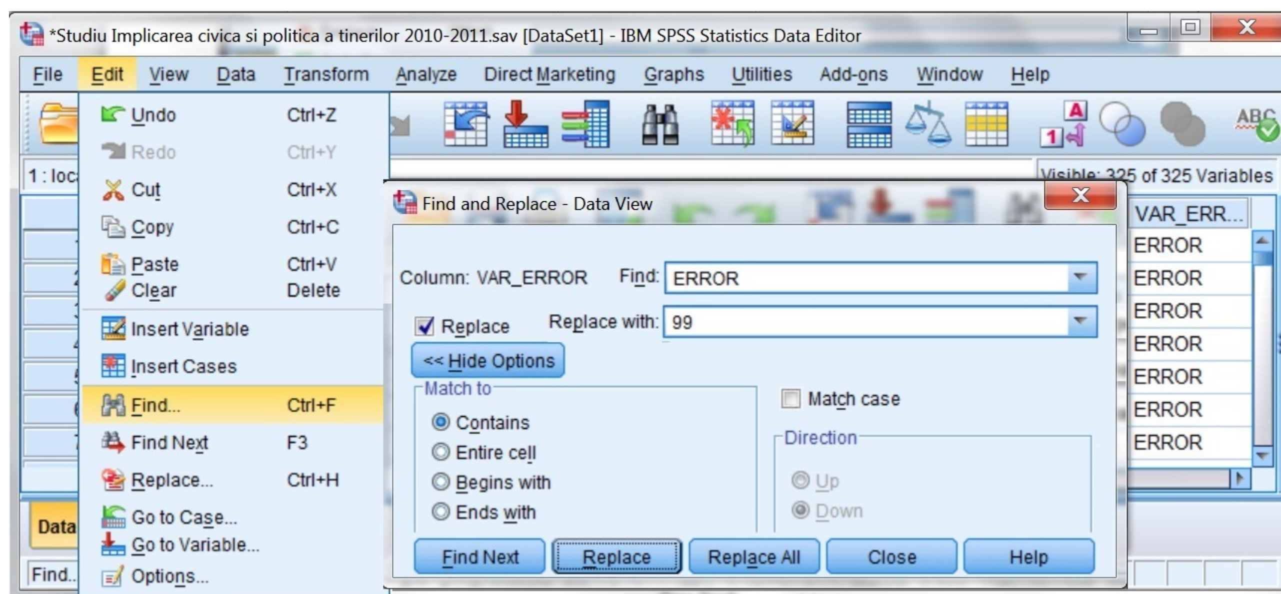
În SPSS ca și în PSPP meniul Edit dispune și de comanda **INSERT VARIABLE**. Avem nevoie de acestă comandă atunci când avem, deja, o bază de date și dorim să poziționăm, între anumite două variabile, una sau mai multe caracteristici noi (aflate, foarte probabil, în relație logică cu cele dintâi). În primul rând se va face spațiu prin una sau mai multe comenzi „Edit Insert Variable", după care în locul nou creat se vor transfera informații prin instrucțiunile discutate anterior („Copy" sau „Cut" și apoi „Paste"). **INSERT CASES** poate viza situația în care indivizii - unitățile statistice de pe liniile bazei de date - au fost introduşi într-o anumită ordine (de exemplu, cea crescătoare după numărul chestionarului) și s-a constatat o omisiune (SPSS Survival manual A step by step guide to data analysis).

În această situație selectăm o linie a bazei de date și apoi dăm comanda Insert Cases; astfel, în baza de date se va face loc pentru introducerea informațiilor despre un nou individ (unitate statistică). Evident, procesul poate fi reiterat dacă e nevoie de introducerea mai multor cazuri.

4.2 Opțiunile EDIT, Find, Go to (case, variable), Replace, în SPSS și PSPP

Instrucțiunea **EDIT FIND** se poate utiliza atunci când căutăm o anumită valoare într-o coloană a bazei de date (reprezentând o caracteristică definită pe populația de referință), indiferent dacă este numerică sau șir de caractere. Atenție: întâi este necesar să ne poziționăm în interiorul coloanei.

Graficul 4.2 Găsirea și / sau modificarea valorilor din baza de date în SPSS



În graficul 4.2 se vede în fereastra de dialog că în coloana VAR_ERROR este căutat cuvântul ERROR.

De asemenea, instrucțiunea **FIND REPLACE** permite găsirea unei valori specifice dintr-o coloana (variabilă) a bazei de date pe care o vizăm și înlocuirea acesteia cu alta (de exemplu se pot înlocui virgulele zecimale care nu sunt acceptate de SPSS pentru valorile numerice cu punctele zecimale). În exemplul din graficul 4.2 am cerut programului să identifice valorile alfanumerice din coloana denumită VAR_ERROR care iau chiar valoarea ERROR și să le înlocuiască cu simbolul 99. Înlocuirea se poate face de fiecare dată cînd SPSS identifică cuvîntul ERROR sau putem utiliza opțiunea Replace All, care va duce la înlocuirea conținuturilor tuturor celulelor din respectiva coloană (valorile string ERROR vor deveni 99 – tot string).

De notat faptul că opțiunea „Edit Find Replace" ne permite și să identificăm căsuțele unei coloane care încep cu anumite caractere (Begins with), se sfârșesc cu anumite caractere (Ends with) sau respectivele caractere sunt conținute în celulele coloanei pe care o vizăm (contains). De asemenea, se poate cere programului SPSS să înlocuiască primele, ultimele sau unele caractere conținute de celulele coloanei cu altele, stabilite de noi (IBM SPSS Statistics 20 Core System, User's Guide).

Menționăm că instrucțiunile „Cut”, „Copy”, „Paste” și „Clear” sunt folosite și în cadrul fișierelor Output din SPSS (care conțin prelucrări de date) într-un mod similar celui discutat anterior dar având particularități specifice datorate posibilității de editare a tabelelor și graficelor create în programul SPSS.

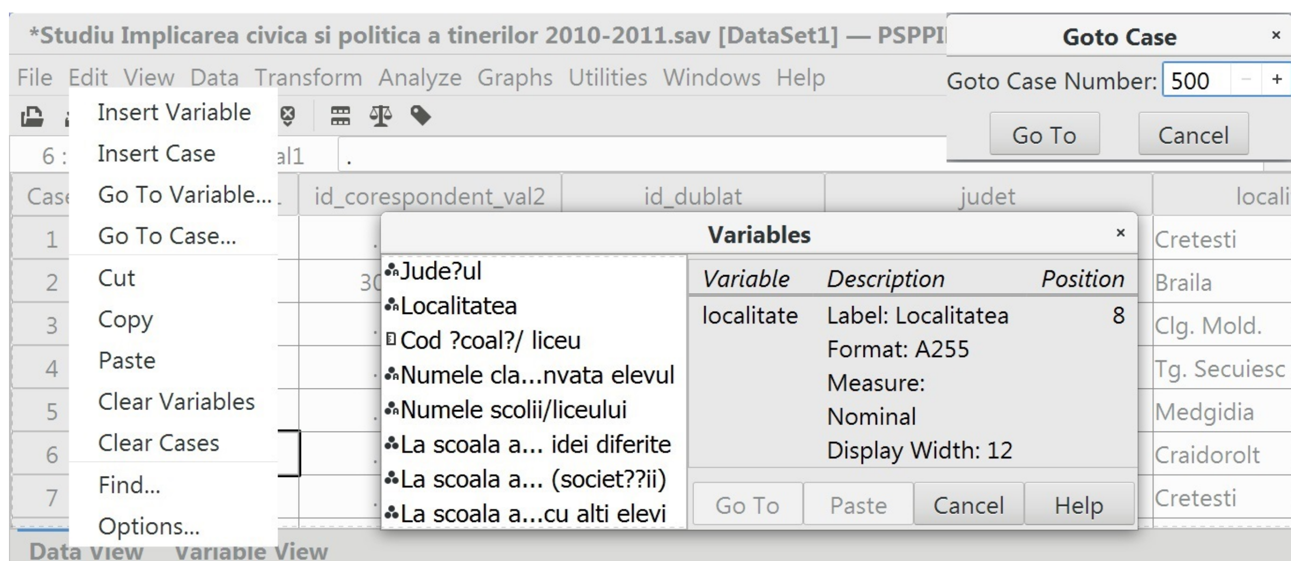
O ultimă observație: utilizând meniul „Edit” din SPSS putem să ne poziționăm în interiorul unei variabile din baza de date dacă îi tastăm numele sau îl selectăm dintr-o listă cu toate variabilele. De asemenea, putem să identificăm o anumită linie a bazei de date, adică un anumit caz, dacă, după succesiunea **EDIT Go To Case**, tastăm indicele corespunzător numărului de ordine care apare în exteriorul părții din dreapta a bazei de date.

Meniul „Edit” oferă o serie de facilități și în programul **PSPP**, chiar dacă există unele diferențe față de SPSS.

Instrucțiunile **DATA Insert Variable** și **DATA Insert Case** sunt similare cu cele din SPSS, fiind utilizate pentru a introduce noi linii (cazuri, unități statistice) sau noi coloane (variabile) în baza de date (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

Instrucțiunea **EDIT Go To Variable** permite selectarea unei caracteristici din lista aflată în stânga ferestrei de dialog (păstrând ordinea din baza de date) și obținerea unei descrieri a respectivei variabile. Dând comanda finală „Go To”, vom fi plasați într-una din căsuțele respectivei caracteristici (SPSS Tutorials, SPSS Beginners Tutorials).

Graficul 4.3 Comanda EDIT Go To Variable în PSPP

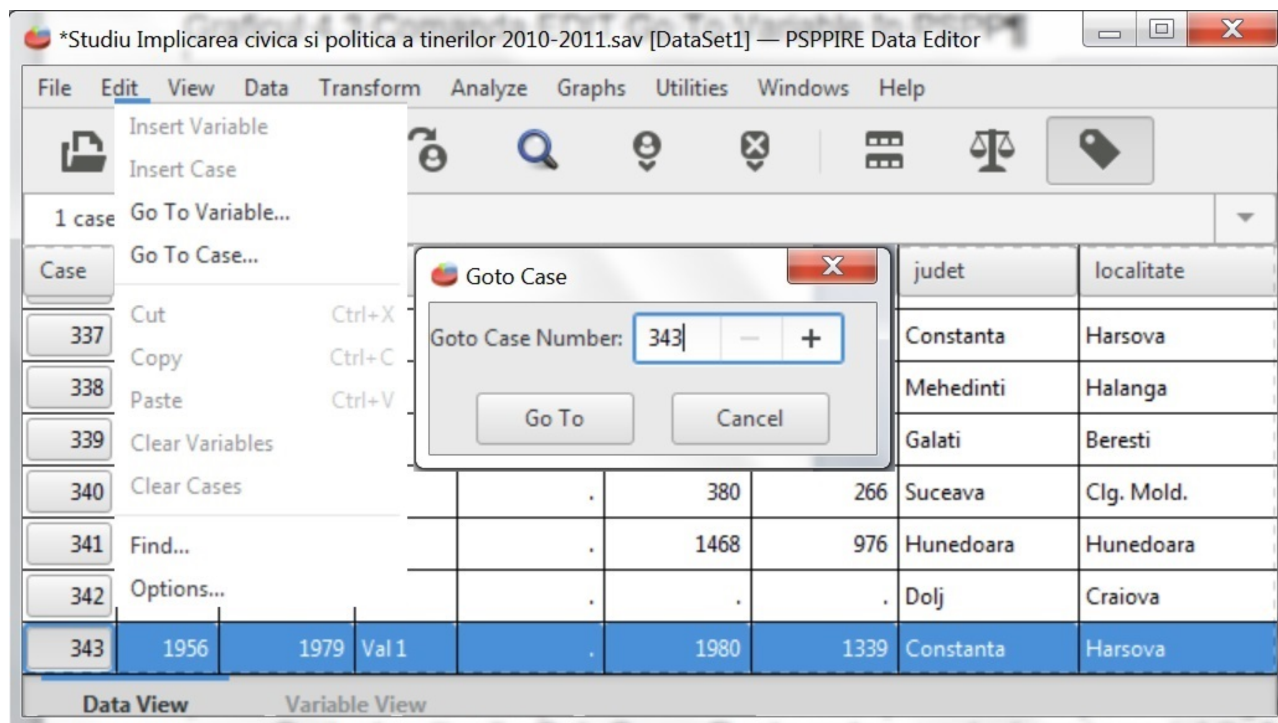


Instrucțiunea **Go To Case** ne permite ca, indiferent unde ne-am situa în baza de date, să putem identifica un anumit caz al acesteia (seria de valori luate de toate variabilele pentru un individ / unitate statistică) tastând numărul cazului (liniei din baza de date) care ne interesează și dând click pe Go To.

Cu instrucțiunile **Cut, Copy**, **Paste** putem muta / copia variabile în baza de date sau putem muta / copia părți din conținuturile unor coloane. Cu instrucțiunea **Edit Clear Variables** pot fi șterse variabile (caracteristici) ale bazei de date iar cu comanda Clear Cases se pot elimina unități statistice, adică se șterg toate datele despre anumiți indivizi conținute în variabile. (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1)

În graficul 4.4 e prezentată seccesiunea ferestrelor care apar atunci când dorim să identificăm un anumit caz din baza de date (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

Graficul 4.4 Comanda Edit Go To Case în PSPP



Este important de reținut că numerele din partea stângă a graficului, care desemnează ordinea cazurilor (unităților statistice) din baza de date nu sunt fixate definitiv. Ordinea poate fi modificată prin acțiuni de sortare a bazei de date folosind drept criteriu de ordonare (crescătoare sau descrescătoare) una sau mai multe din variabilele existente (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

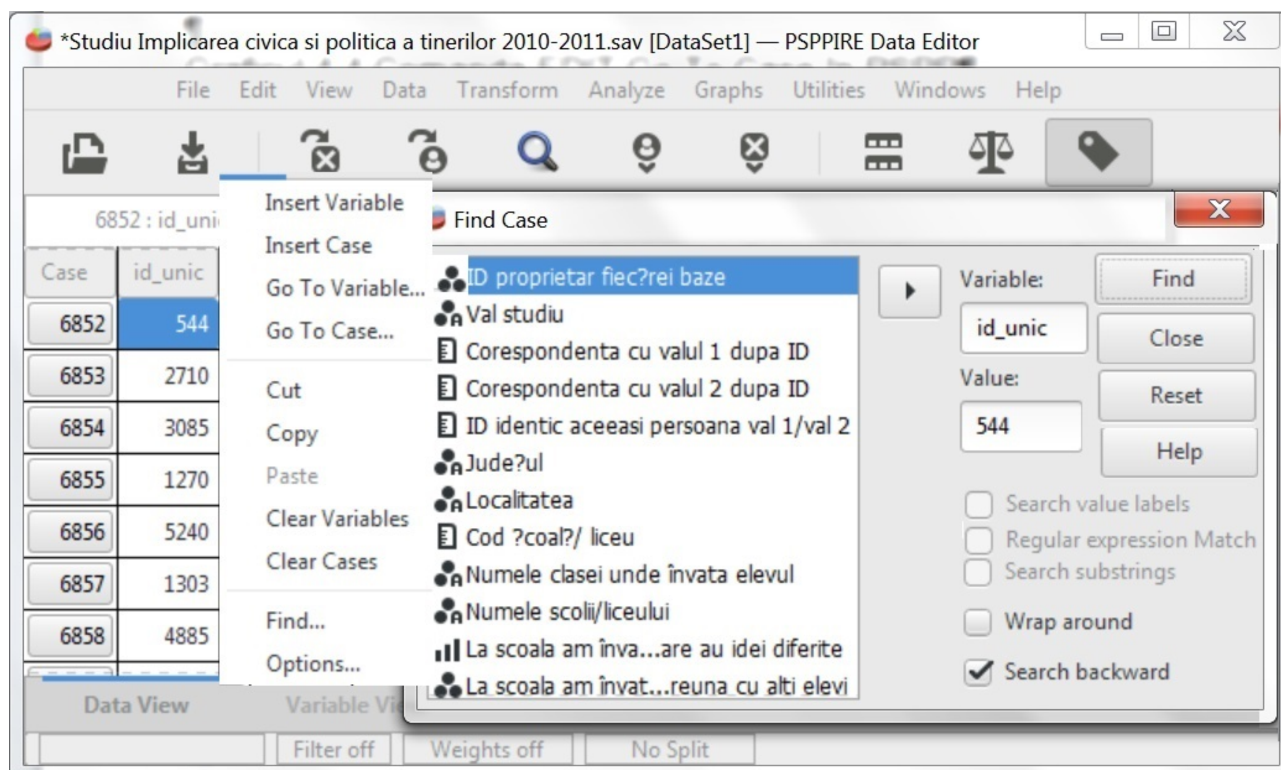
În graficul 4.5 este indicat modul în care putem identifica o valoare a unei variabile (caracteristici) definită în baza de date prin intermediul comenzilor **EDIT Find** în programul PSPP.

În momentul în care dăm succesiunea de comenzi „Edit Find", va apărea o fereastră de dialog în care, în partea stângă, vom avea lista variabilelor din baza de date. Din această listă o vom selecta variabila care ne interesează dând click și apoi o vom introduce în căsuța numită „Variable" din dreapta ferestrei (cu ajutorul săgeții din dreptul căsuței). Trebuie reținut faptul că, în funcție de modul în care sunt definite variabilele dpdv. sociologic, găsirea valorilor se va realiza în mod diferit. În lista din graficul 4.4 există niște simboluri înaintea numelor varabilelor despre care am mai discutat:

- simbolul pentru „**ID proprietar fiecărei baze**" semnifică o variabilă *calitativă nominală introdusă sub formă numerică*
- simbolul pentru „**Val studiu**" indică o variabilă *calitativă nominală introdusă sub formă de text* (string) în baza de date
- simbolul pentru „**La şcoală am învăţat să-i înţeleg pe oamenii care au idei diferite**" indică o variabilă *calitativă ordinală introdusă numeric*

- simbolul pentru **„Cod Școală / liceu"** indică o *variabilă cantitativă introdusă numeric* (atenție: variabilele string nu pot fi declarate cantitative)

Graficul 4.5 Comanda Edit Find în PSPP



În graficul 4.5 a fost selectată variabila „ID proprietar fiecărei baze" și introdusă în căsuța „Variable" (unde apare sub forma numelui variabilei, care este „id_unic"). În acest caz sunt active două forme de căutare pentru valoarea pe care vrem să o găsim printre valorile lui „id_unic": *Wrap around* respectiv *Search backward*. Nu am reușit să identificăm diferențe între cele două variante, ambele ne conduc la valoarea variabilei dacă am introdus o dată validă (în caz contrar nu este furnizat nici un rezultat).

Dacă selectăm variabila „Val studiu", care este calitativă, introdusă sub forma string (șir de caractere) în baza de date și pentru care s-au introdus și etichete ale valorilor, toate opțiunile de căutare sunt active.

Dacă se alege „Search value labels" și se indică o etichetă validă, dând comanda „Find" vom fi poziționați în coloana variabilei „Val studiu" pe un caz având eticheta solicitată.

Dacă alegem „Regular expression Match" va trebui să alegem una dintre valorile variabilei (și nu etichete ale acesteia) pe care să le introducem în căsuța „Value:" Dacă valoarea este validă, vom fi conduși la un caz având respectiva valoare (în caz contrar nu se înregistrează nici o schimbare).

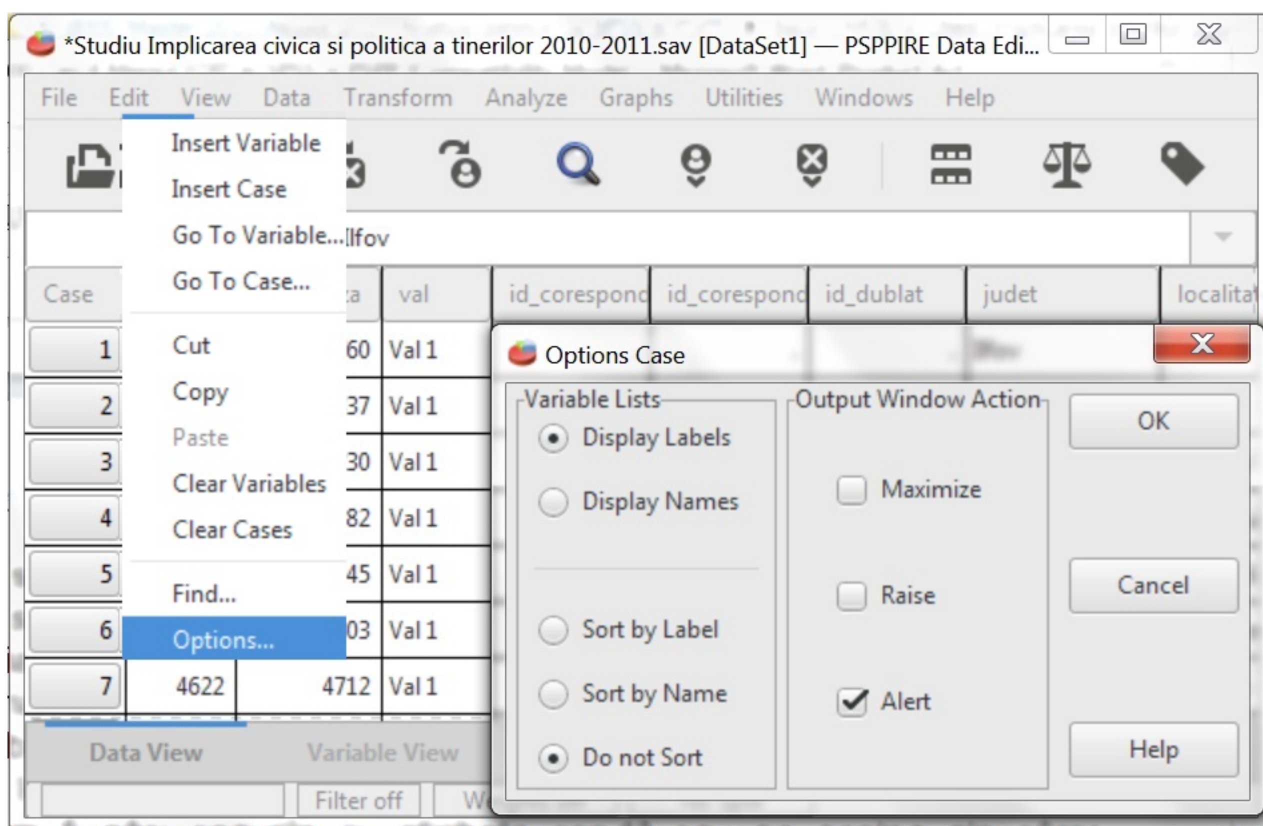
Dacă se alege „Regular expression Match" sau „Search substrings", în căsuța „Value" se poate introduce expresia completă a valorii variabilei „string" sau numai o succesiune dintre literele care o compun. În primul caz prin comanda „Find" ne va conduce la valoarea exactă pe care am indicat-o. În a doua situație, în variabila „Val studiu" vor fi identificate toate căsuțele pentru care se înregistrează potriviri între valorile conținute și căutarea care a fost lansată (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.)

Dacă lansăm căutarea după o anumită variabilă, putem identifica atât valorile acesteia cât și eticheta (dacă am ales „Search variable labels”). În această din urmă situație, dacă optăm și pentru „Search substrings”, vom putea să selectăm cazurile variabilei indicând o etichetă în întregime sau o succesiune de caractere din conținutul ei. Ca rezultat, vom fi găsi cazurile cu eticheta cerută sau pe acelea ale căror etichete conțin literele indicate.

4.3 Varianta EDIT, Options, în PSPP

Ultima variantă a meniului **EDIT**, în PSPP, este „**Options**”. În partea din stânga a ferestrei de dialog există câteva aspecte care pot fi setate privitor la lista variabilelor. Se poate alege ca lista să conțină etichetele variabilelor – DisplayLabels - sau numele acestora - Display names (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

Graficul 4.6 Comanda Edit Options în PSPP



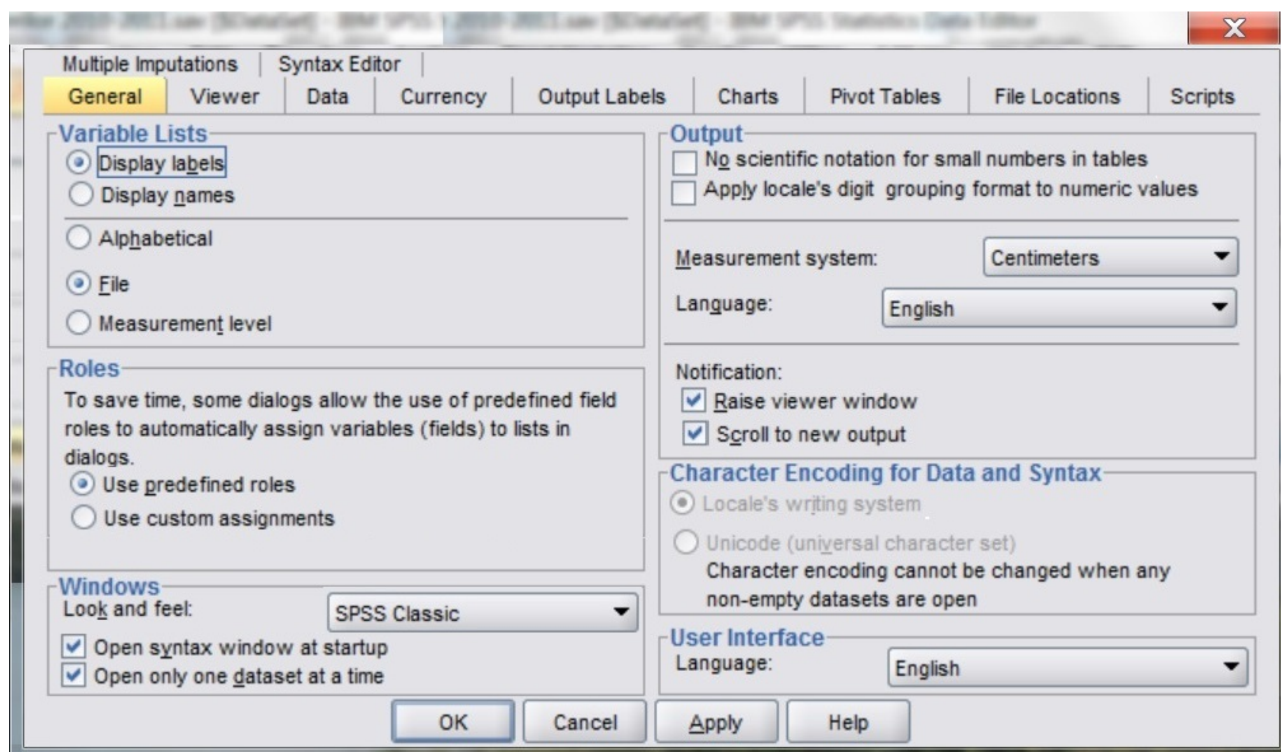
În continuare, se poate opta pentru prezentarea variabilelor în ordinea alfabetică a etichetelor („Sort by Label”), în ordinea numelor variabilelor („Sort by Name”) sau se poate alege varianta „Do not Sort”, caz în care ordinea listei variabilelor va fi identică cu cea existentă în baza de date.

În partea centrală a ferestrei de dialog există câtva opțiuni care se referă la prezentarea ferestrei Output (în care sunt stocate rezultatele analizei datelor și mesajele care se doresc a fi transmise utilizatorilor programului PSPP). Selectarea opțiunii „Alert” permite celor ce folosesc PSPP să obțină informații despre pașii urmați de program iar „Maximize” și „Raise” vizează, în principal, dimensiunea și prezentarea tabelelor și graficelor (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

Vom continua descriind aspectele principale ale opțiunii „Edit Options” dar, de această, dată, pentru programul SPSS (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

4.4 Opțiunea EDIT, Options, General, Viewer, Data, în SPSS

Graficul 4.7 Comanda Edit Options General în SPSS



În partea din stânga a ferestrei de dialog obținută prin comanda **Edit Options General** se poate stabili dacă în lista de variabile prezentată în comunicarea cu utilizatorii programului SPSS vor fi prezentate etichetele variabilelor sau numele respectivelor variabile.

Mai jos, se poate alege modul în care să fie listate variabilele: se poate opta pentru ordonarea alfabetică, pentru păstrarea succesiunii existente în baza de date (opțiunea File) sau se poate folosi drept criteriu nivelul de măsurare al variabilei: întâi se vor plasa variabilele calitative nominale, apoi cele calitative ordinale și, în final, cele cantitative, fiecare dintre categorii fiind ordonată alfabetic (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

În dreapta ferestrei de dialog se poate alege folosirea sau nu a notației științifice pentru numerele mici din tabele (pentru 0.000047 se poate folosi notatia stiintifica 4.7E-005, semnificând 4.7 înmulțit cu 10 la puterea -5).

În continuare se poate alege unitatea de măsură care să fie utilizată în dialogurile cu utilizatorii SPSS și limba în care să aibă loc acest dialog.

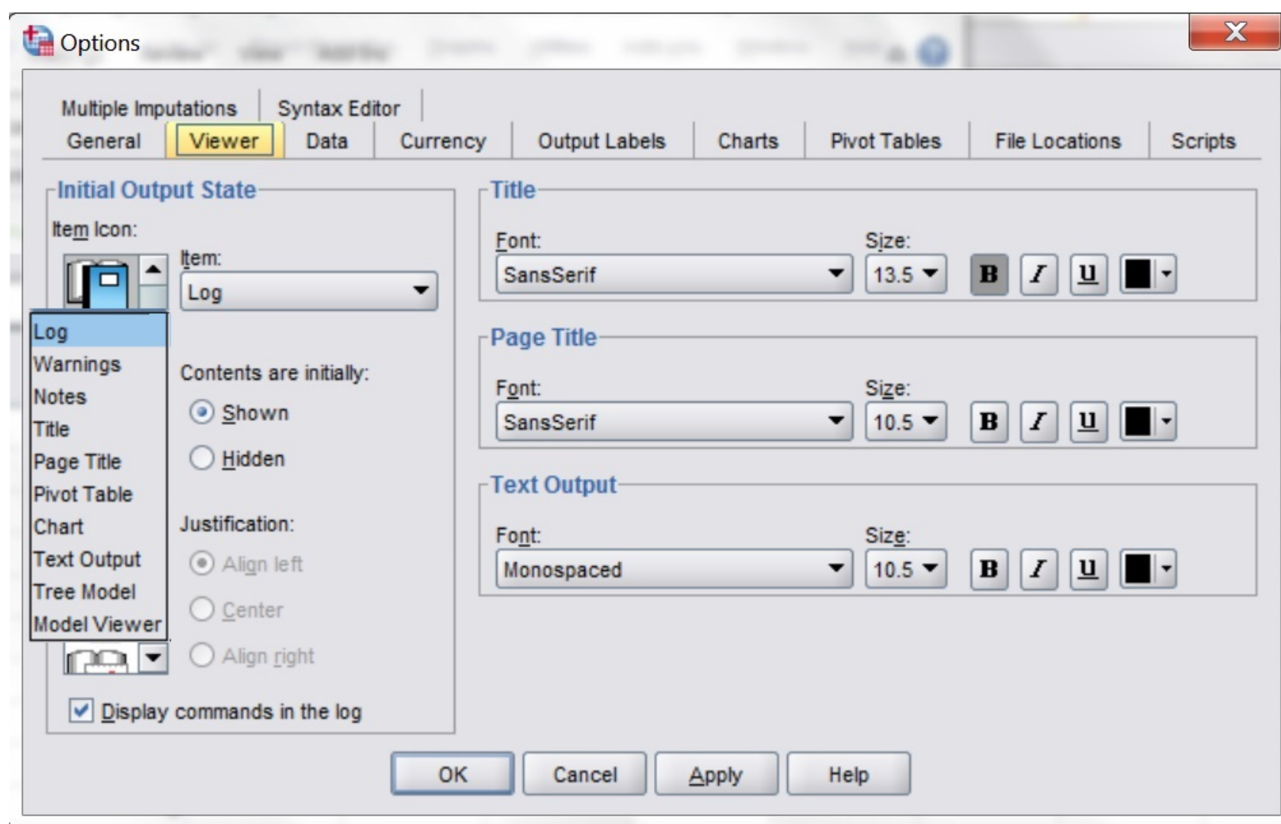
Opțiunile „Scroll to new output” și „Raise viewer window” se referă la modul în care se operează cu cu obiectele output create prin analiza datelor. Putem exemplifica referindu-ne la realizarea unor tabele, lansarea editării unuia dintre ele - dând prin dublu click deasupra acestuia - și apoi crearea unui nou tabel, fapt ce poate determina inactivarea celui dintâi sau păstrarea lui în starea de editare (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Opțiunea „Character encoding for data and Syntax" nu este activă iar ultima alegere pe care o poate face utilizatorul în această fereastră de dialog este selectarea interfeței (limba română nu este printre variante).

Vom continua cu descrierea (parțială) a ferestrei de dialog obținută prin succesiunea de comenzi **Edit Options Viewer** care se referă la vizualizarea diferitelor elemente legate de fișierele output (informații și prelucrări de date având extensia .spv).

În primul rând, în partea stîngă a ferestrei există posibilitatea alegerii itemului la care se vor referi opțiunile de activare / deazactivare respectiv de aliniere (stânga, dreapta sau centru, acolo unde e cazul). În partea cea mai de jos a graficului 4.8 putem alege activarea opțiunii „Dispaly commnands în the log" ceea ce înseamnă că in fereastra „Log" din Output vor apărea comenzile scrise in mediul de programare SPSS care au fost utilizate (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 4.8 Comanda Edit Options Viewer în SPSS



Dacă „Display commands in the log" este selectat atunci vizualizarea sau ascunderea itemului „Log" corespunde prezenței sau absenței, în fișierul output, a setului de comenzi SPSS anterior menționate (respectivul set poate fi folosit, într-un fișier de tip sintaxă, pentru a relua automat pașii anterior generați prin succesiuni de comenzi date utilizând ferestele de dialog SPSS).

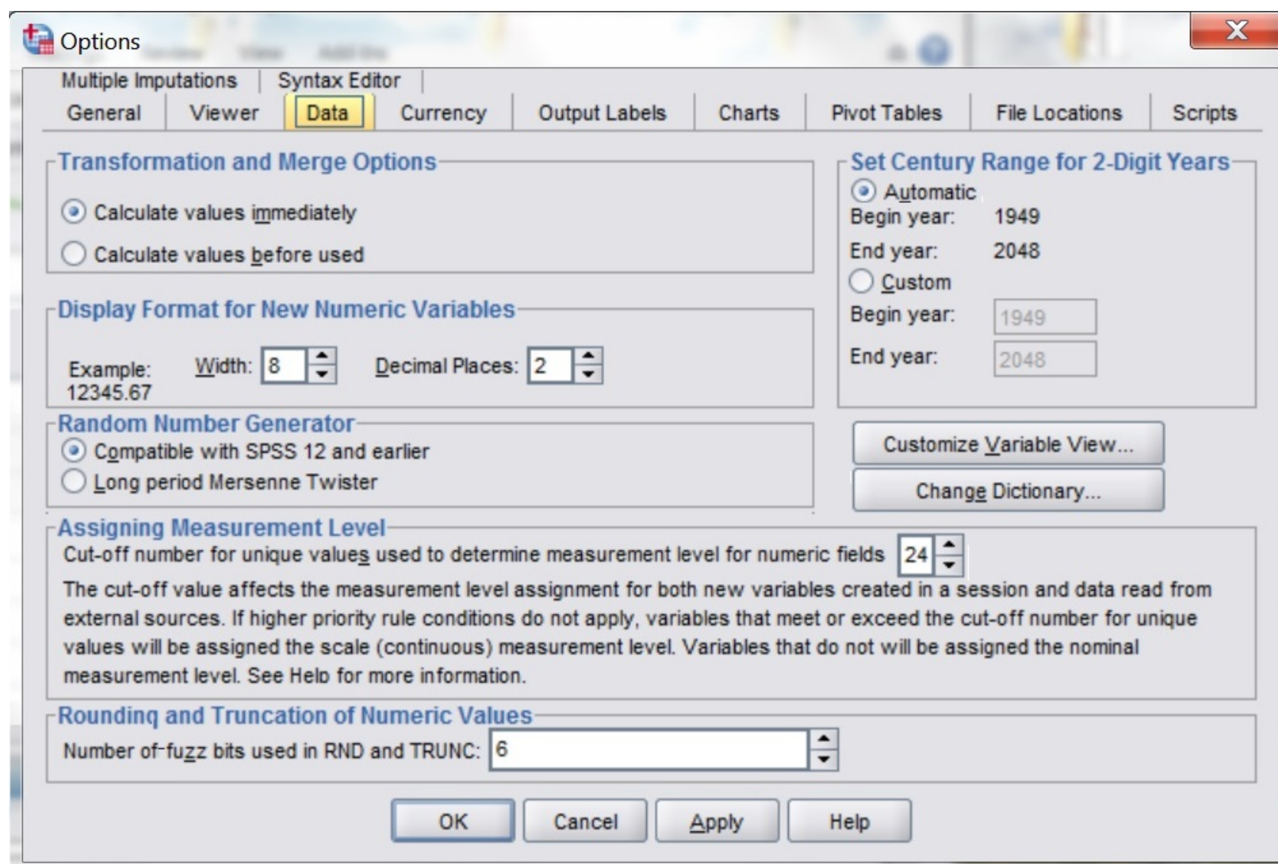
Partea stângă a ferestrei se referă, în plus, la activarea, ascunderea și elimnierea (unde e cazul) a următoarelor elemente prezente în fișierul output: Avertismente și Note (informații date de SPSS), prezența / absența, în output a titlului unui tabel sau a titlului paginii, vizualizarea și aranjarea în pagină a tabelelor sau a graficelor, prezentarea secvențelor de text, etc. (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

În partea dreaptă a ferestrei se pot seta, în primul rând, „Titlul” respectiv „Titlul paginii” (corespunzând unei singure comenzi SPSS, chiar dacă are ca rezultat obținerea mai multor tabele sau grafice). Se pot alege, de asemenea, tipurile de font, dimensiunile acestora și alte caracteristici, inclusiv culoarea.

Pentru secvențele de text, cum ar fi cele care indică variabilele care au fost cuprinse într-o anumit set de analize statistice, opțiunile sunt asemănătoare: tipul de font, mărimeaa acestuia, faptul de a fi îngroșat sau normal, înclinat sau obișnuit, subliniat sau nu, respectiv culoarea fontului.

Ne vom referi, în cele ce urmează, la comanda **Edit Options Data**, care reglementează, în principal, aspectele privitoare la calculul / tratamentul datelor în programul SPSS (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 4.9 Comanda Edit Options Data în SPSS



Pentru căsuța „Transformation and marge options” se poate seta dacă o serie de valori sau chiar baze de date să fie calculate (construite) în momentul în care a fost dată comanda în SPSS sau atunci când se face prima oară apel la respectivele date sau structuri de date. De exemplu, variabila desemnând media generală a anului trecut școlar poate fi recodificată în mai multe categorii. Calculul noii variabile se poate face de la început (după comanda Trnsforme Recode) sau în momentul in care se dorește pentru prima oară folosirea valorii variabilei definită pe categorii de reușită școlară.

În căsuța „Display format for new numeric variables” apare forma implicită a variabilelor numerice nou create (în exemplul din graficul 4.9 apar 2 zecimale, punctul zecimal și maximum 5 cifre semnificative) (IBM Knowledge Center, Home, SPSS Statistics 24.0.0)

Căsuța „Random Number Generator” permite selectatea algoritmului de generare a numerelor aleatoare, putându-se alege între varianta mai veche, valabilă pînă la SPSS 12 și cea mai nouă (Mersenne Twister Random Number Generator. Vezi help-ul programului SPSS, „IBM Knowledge Center, Home, SPSS Statistics 24.0.0”

În căsuța „Assigning Measurement Level” se stabilește de la ce număr de valori diferite luate de variabilele nou create sau citite dintr-o sursă externă acestea să fie descrise ca fiind continue (Scale) respectiv calitative (Nominal). Criteriul îl reprezintă acel „Cut off number for unique values” folosit pentru a determina tipul de scală de măsură pentru variabilele introduse numeric în baza de date.

Căsuța „Rounding and truncation of numeric Values” se referă la situațiile în care este nevoie de setarea numărului de biți prin care o valoare calculată sau introdusă poate diferi ușor de o limită de rotunjire sau trunchiere și să fie considerată în continuare egală cu limita respectivă minus unele „fuzz” sau erori.

Pentru funcțiile RND și TRUNC, căsuța „Rounding and truncation of numeric Values” controlează pragul implicit pentru rotunjirea valorilor care se află foarte aproape de o limită de rotunjire. A se vedea „IBM Knowledge Center, Home, SPSS Statistics 24.0.0”

Căsuța „Set Century Range for 2-Digit Years” definește intervalul de ani pentru variabile dată calendaristică introduse / afișate cu anul reprezentat prin 2 cifre (ex., 10/28/86, 29-OCT-87). Setarea automată a intervalului se bazează pe anul curent, începând cu 69 de ani înainte și 30 de ani după anul în curs (adăugând anul curent, avem un interval de 100 de ani). Pentru un interval personalizat, anul final e determinat automat pe baza valorii introdusă pentru anul de început (IBM Knowledge Center, Home, SPSS Statistics 24.0.0)

Căsuța „Customize Variabile View” permite stabilirea elementelor care vor fi prezente în modul de prezentarea „Variable view” a bazei de date: „Name” (numele variabilei), „Type” (numeric, string, etc.), „Width” (număr de caractere alocat), „Decimals” (număr maxim de zecimale acceptat), „Label” (eticheta numelui variabilei), „Values” (etichetele valorilor variabilei), „Missing” (datele definite ca fiind lipsă), „Columns” (numărul de caractere vizibil din coloana variabilei), „Align” (alinierea valorilor variabilei la stânga, dreapt, mijloc), „Measure” (nominal, ordinal sau cantitativ – scale), „Role” (independentă, dependentă, de ambelel tipuri, nici una din variante) respectiv „Partition” și „Split” (referindu-se la utilizarea variabilelor pentru divizarea bazei de date în mai multe seturi, utilizabile independent).

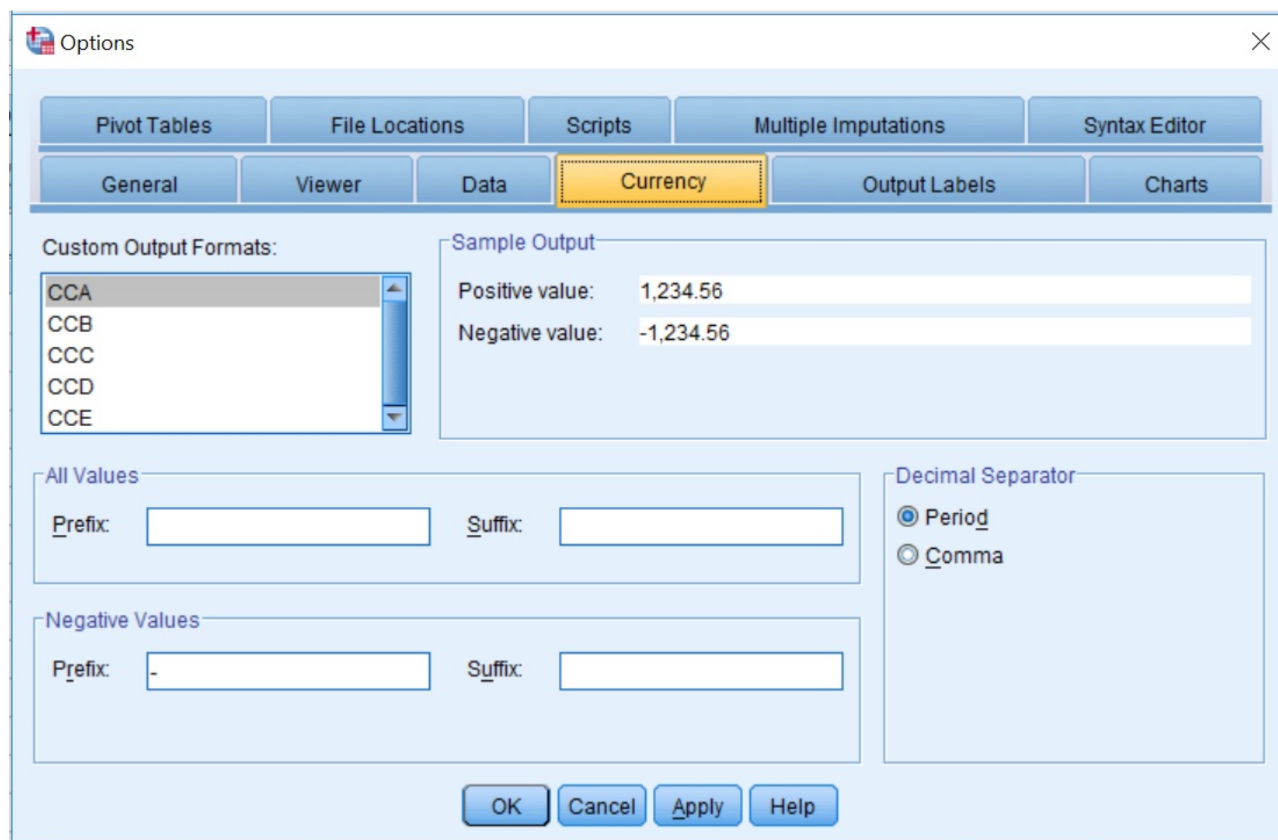
Căsuța „Change Dictionary” permite să se aleagă, în programul SPSS, unul dintre dicționarele care sunt disponibile pentru a fi folosite la verificarea ortografică a elementelor din „Variable View”.

Comanda „Edit Options Currency”, pe care o prezentăm în continuare, se referă la definirea unor formate particulare de prezentare a datelor.

Fereastra de dialog permite definirea a 5 formate particulare, CCA, CCB, CCC, CCD, și CCE, prin alegerea prefixelor și sufixelor pentru toate numerele respectiv pentru cele negative. Din stânga ferestrei de dialog se poate alege separatorul zecimal punct („Period”) sau virgulă („Comma”).

În exemplul din graficul 4.10 este activă opțiunea CCA pentru care s-a setat prefixul „-” pentru numerele negative și separatorul zecimal „Period” (punct). Nu au fost setate prefixe / sufixe pentru toate valorile respectiv sufixe pentru cele negative (virgula va marca miile, milioanele, etc.)

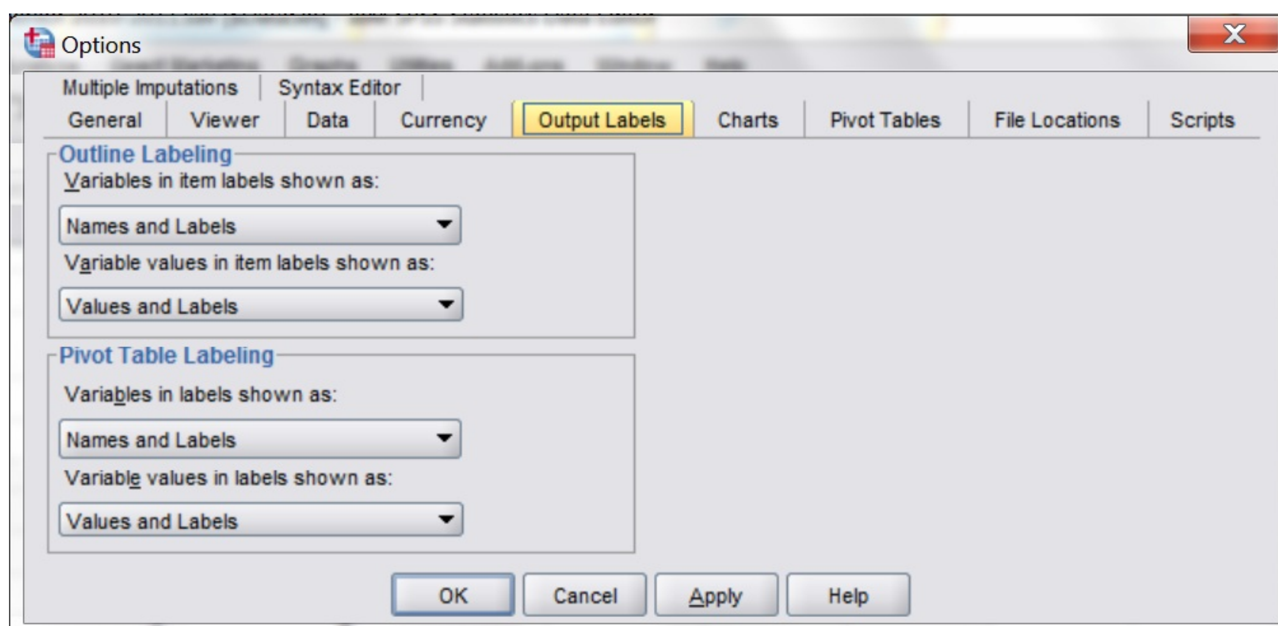
Graficul 4.10 Comanda Edit Options Currency în SPSS



4.5 Varianta EDIT, Options, Output Labels, Charts, Pivot Tables, în SPSS

Următoarea comandă, **Edit Options Outpult Labels**, se referă atât la etichetele variabilelor și ale valorilor acestora care apar in fisierele de tip „Output" sub formă de titluri sau seturi de informații oferite de programul SPSS precum și la etichetele variabilelor și valorilor din interiorul tabelelor generate de program (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 4.11 Comanda Edit Options Output Labels în SPSS

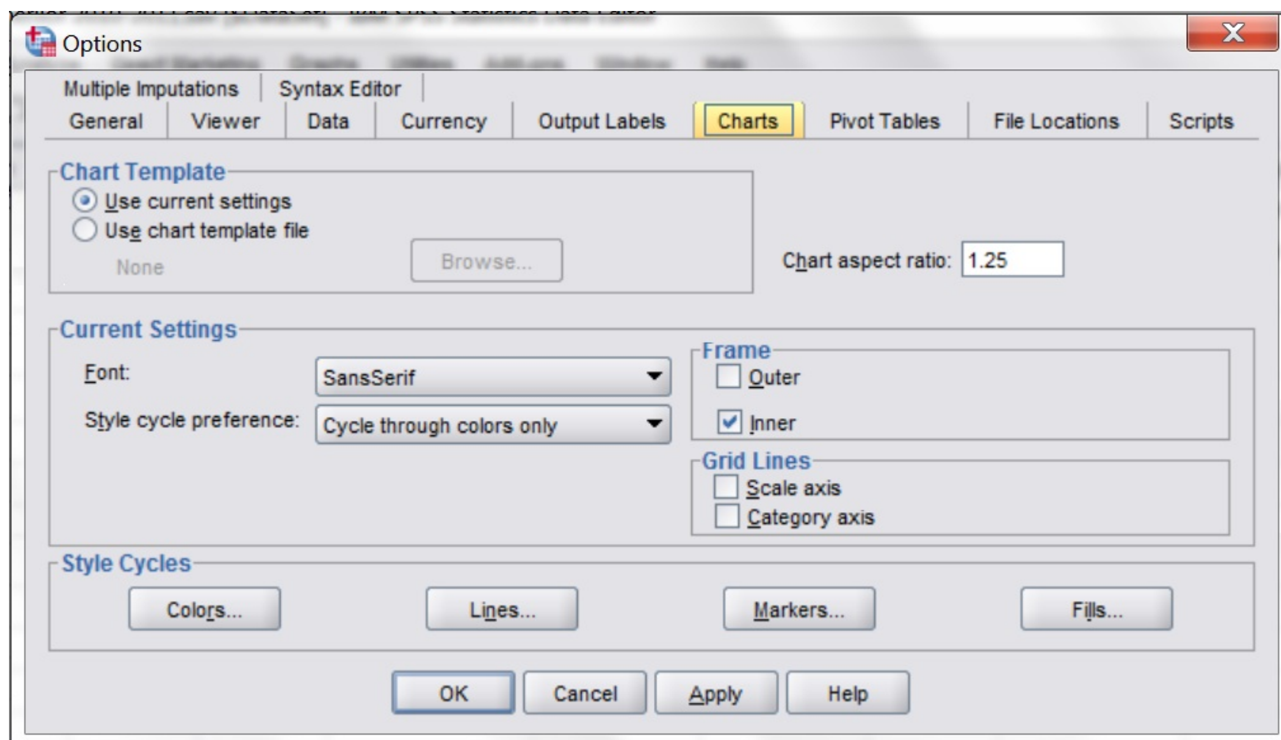


Dacă alegem „Outline Labeling” putem stabili dacă informațiile sau mesajele care apar în fișierele output (în exteriorul tabelelor generate de programul SPSS) vor conține nume, etichete sau nume și etichete ale *variabilelor* respectiv valori, etichete sau valori și etichete ale *valorilor variabilelor* (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Dacă utilizăm opțunea „Pivot Table Labeling” vom putea seta aceleași chestiuni (vizualizarea numelul variabilei și a valorilor) dar referitor la interiorul tabelelor generate de programul SPSS (se vor putea alege nume, etichete sau nume și etichete ale *variabilelor* respectiv valori, etichete sau valori și etichete ale *valorilor variabilelor*).

În graficul 4.12 este prezentată fereastra de dialog obținută prin comenzile **Edit Options Charts** în SPSS.

Graficul 4.12 Comanda Edit Options Charts în SPSS



În partea de sus din stânga a ferestrei de dialog se poate alege între utilizarea setărilor curente (Use current settings) și folosirea unui format predefinit pentru grafic, având extensia .sgt (Use chart template file).

Dacă optăm pentru setările curente, putem alege, din stânga ferestrei, fonturile care să fie folosite pentru valorile din interiorul graficelor (Inner frame) sau în proximitate (Data Frame respectiv Legend Frame).

„Style cicle preferance” se referă la atribuirea inițială a culorilor și modelelor de fond pentru graficele nou-construite.

„Cycle through colors only” utilizează doar culori pentru a diferenția elementele diagramei și nu utilizează modele.

„Cycle through patterns only” utilizează numai stiluri de linie, simboluri de marcare sau modele de umplere a diferitelor suprafețe pentru a diferenția elementele diagramei și nu utilizează culoarea (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

„Frame Outer sau Inner” controlează afișarea fonturilor și a culorilor, a stilurilor de linie, a simbolurilor de marcare sau a modelelor de umplere numai pentru elementele interioare ale graficului, numai pentru cele exterioare sau pentru ambele variante în cazul diagramelor nou-create (IBM Knowledge Center, Home, SPSS Statistics 24.0.0)

„Chart Aspect Ratio” reprezintă raportul dintre lățimea și înălțimea cadrului exterior al noilor grafice. Se poate specifica un raport lățime-înălțime începând cu 0,1 și mergând până la 10,0. Valorile subunitare generează grafice mai înalte și mai înguste.

Valorile supraunitare generează diagrame cu lățime mai mare decât înălțimea iar o valoare unitară produce diagrame pătratice. Este important de reținut faptul că după ce o diagramă a fost creată, raportul dintre înălțime și lățime nu mai poate fi schimbat (IBM Knowledge Center, Home, SPSS Statistics 24.0.0) .

„Grid Lines” permite, în interiorul graficului, selectarea vizualizării liniilor paralele cu axa orizontală, axă definită drept scală de măsurare a variabilei (Scale axis) iar selectarea opțiunii „Cathegory axis” dă posibilitatea vizualizării liniilor verticale, paralele cu axa frecvențelor categoriilor variabilie (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

„Style cycles” oferă posibilitatea, pentru graficele ce vor fi construite în continuare, de a alege diferitele culori interioare („Colors” – dacă s-a optat anterior pentru varanta „Cycle through colors only”) sau a patternurilor de umplere a elementelor interioare graficului („Fills” dacă s-a optat anterior pentru „Cycle through patterns only”).

„Lines” și „Markers” sunt două opțiuni vizând construcția graficelor, prima referindu-se la tipul de linie care să fie folosită (continuă, punctată, etc.) iar a doua, la formele geometrice care să fie utilizate în interiorul diagramelor pentru a evidenția anumite proprietăți.

4.6 Varianta EDIT, Options, Pivot Tables, Sintax Editor, Scripts, în SPSS

Graficul 4.13 ilustrează fereastra de dialog care e deschisă, în programul SPSS, prin succesiunea de comanzi **Edit Options Pivot Tables**.

În căsuța „Table look” se poate opta între o serie de modalități de prezentare a tabelelor de asociere, prin simpla selecție a uneia dintre variante (care va fi folosită în procedura „Analyze, Descriptive Statistics, Crosstabs”).

Sub căsuță „Table look” există opțiunea „Browse” care permite indicarea căii de acces spre oricare fișier având structura unui format de tabel de asociere (extensia .stt). Butonul „Set Table Look Directory” este utilizat pentru a seta folderul în care programul SPSS prezintă formatele implicite ale tabelelelor de asociere.

Dacă dorim ca, în mod implicit, în „Table Look” să apare fișierele care descriu structura tabelară (având extensia .stt) din alt folder, este suficient să selectăm un format din respectivul folder (utilizând „Brows”), să optăm pentru „Ok” și apoi să dăm click pe butonul intitulat „Set Table Look Directory”

Căsuța „Column Widths” setează ajustarea automată a lățimii coloanelor în tabele de asociere. Prima opțiune, „Adjust for labels and data except for extremely large tables” vizează ajustarea etichetelor și datelor, cu excepția tabelelor foarte mari.

Pentru tabelele sub 10.000 de celule, se ajustează lățimea coloanei la maximum dintre numărul de caractere al eticheta coloanei și cea mai mare valoare a datelor (IBM SPSS Statistics 20 Core System, User’s Guide).

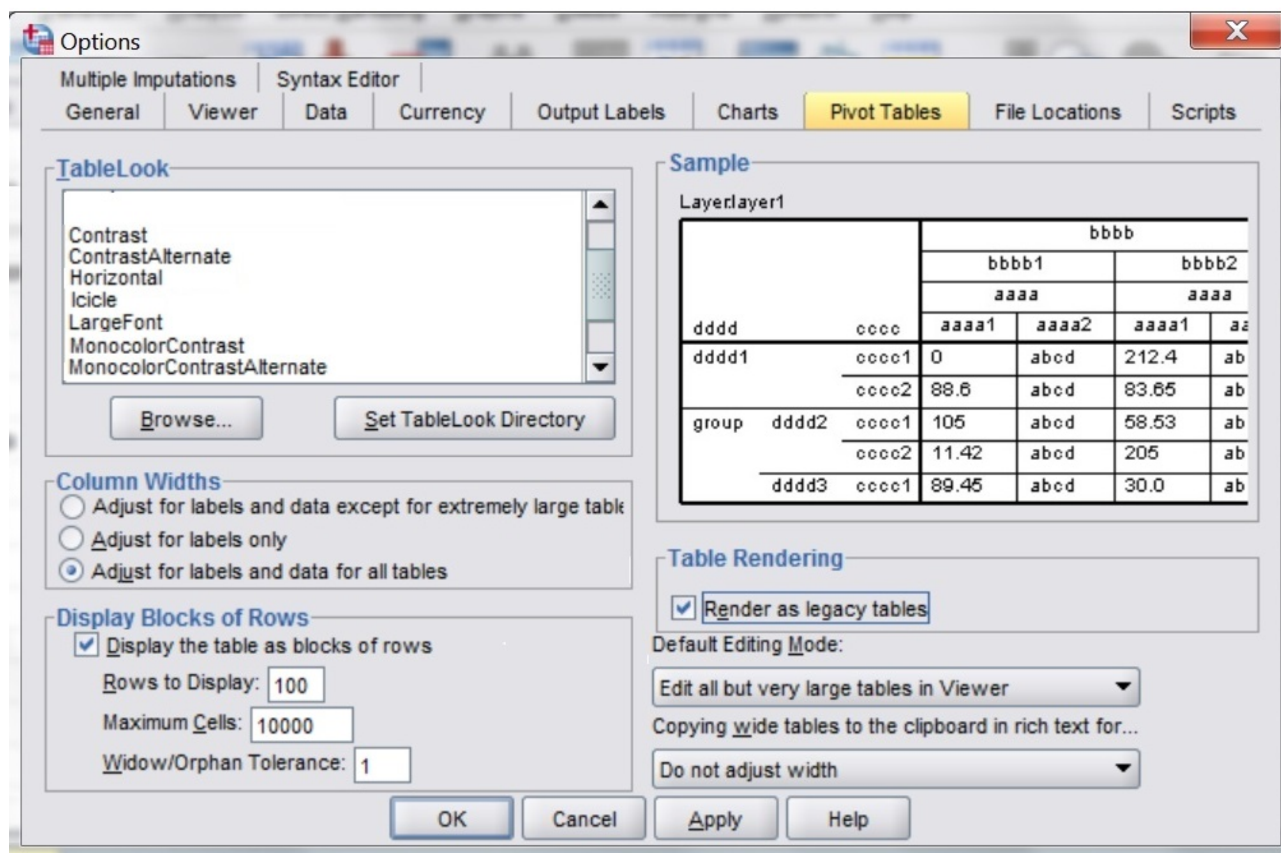
Pentru tabelele peste 10.000 de celule, se ajustează lățimea coloanei la cea a etichetei coloanei. (Notă: opțiune e disponibilă doar dacă se selectează „Render as Legacy Tables".)

În aceeași căsuță de dialog „Column Widt" putem alege „Adjust for labels only", ceea ce înseamnă că lățimea coloanei va fi ajustată la lățimea etichetei coloanei (prin această procedură se vor obține tabele mai compacte dar valorile cu mai multe caractere decât cele din eticheta variabilei vor fi trunchiate).

Printre opțiunile căsuței de dialog „Column Widt" există și varianta „Adjust for labels and data for all tables".

Lățimea coloanei va fi ajustată la cel mai mare număr de caractere dintre cele aflate în componența etichetei variabilei și maximumul valorilor datelor. În aceste condiții se vor obține tabele cu lațime mai mare dar se asigură afișarea tuturor valorilor netrunchiate (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 4.13 Comanda Edit Options Pivot Tables în SPSS



Opțiunea „Display Blocks of Rows" se referă la afișarea unor tabele mari în fișierul Output. De notat că setările nu influențează tipărirea tabelelor mari. (opțiunea e disponibilă doar dacă selectăm „Render as legacy tables".)

Selectând „Display the table as blocks of rows", tabelele vor fi afișate în fisierul output pe secțiuni, succesiv vizibile folosind comenzile de navigare.

„Rows to display" stabilește numărul de rânduri afișate într-o secțiune. Valoarea este între 10 și 1000.

„Maximum cells" stabilește numărul maxim de celule care va fi afișat în fiecare secțiune. Valoarea trebuie să fie între 1000 și 100000 (IBM SPSS Statistics 20 Core System, User's Guide).

Opțiunea „Widow/orphan tolerance" va fi utilizată când se afișează tabele în output cu un număr fix de rânduri într-o secțiune (așa cum este stabilit prin căsuța de dialog „Rows to display"). Valoarea implicită este zero.

Dacă o pauză între blocurile de rânduri lasă un numă de rânduri „văduve" egal sau mai mic decât toleranța specificată, atunci zona de pauză este mutată mai sus în tabel pentru a se afișa acele rânduri în blocul următor.

Dacă o pauză între blocurile de rânduri lasă un numă de rânduri „orfane" egal sau mai mic decât toleranța specificată, atunci zona de pauză este mutată mai jos în tabel pentru a afișa acele rânduri în blocul următor.

Căsuța de dialog „Default Editing Mode" permite activarea tabelelor în fereastra Viewer sau într-o fereastră separată. Implicit, dublu-clic asupra unui tabel îl activează în fereastra Viewer, cu excepția celor foarte mari.

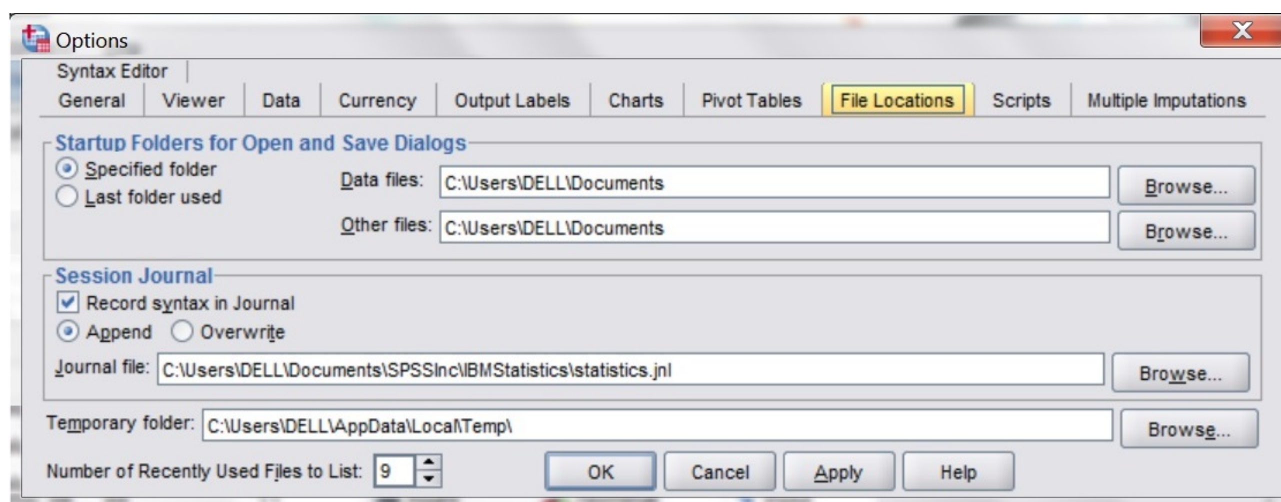
Se poate alege activarea tuturor tabelelor în fereastra Viewer („Edit all tables in Viewer"), într-o fereastră separată („Open all tables in a separate window") sau avem posibilitatea de a opta pentru deschiderea tabelelor în fereastra viewer cu excepția celor foarte mari – „Edit all but very large tables in Viewer" (IBM SPSS Statistics 20 Core System, User's Guide).

Opțiunea „Copying wide tables to the clipboard in rich text format" e utilizată când tabelele sunt copiate în în format Word / RTF.

Tabelele prea mari pentru lățimea documentelor pot fi îngustate pentru a se potrivi lățimii documentului („Shrink width to fit") sau pot fi lăsate neschimbate („Do not adjust widt"). Dacă se alege „Wrap table" tabelul va fi descompus într-o succesiune de tabele cu un număr mai mic de coloane astfel încât să încapă în documentul Word sau RTF.

Prin Comanda „**Edit Options Sintax Editor**" setăm folderul implicit pe care programul îl va folosi pentru a deschide și salva fișierele la începutul fiecărei sesiuni, stabilim poziționarea fișierului jurnal și indicăm locul în care se află folderul temporar precum și numărul de fișiere care apar în lista celor utilizate recent (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 4.14 Comanda Edit Options Sintax Editor în SPSS



Căsuța de dialog „Folders for Open and Save Dialog Boxes" se referă la alegerea folderelor implicite pentru deschiderea și salvarea fișierelor dialog care conțin date de ieșire respectiv care au alte structuri – „Specified folder" (IBM SPSS Statistics 20 Core System, User's Guide).

Alegând varianta „Last folder used”, deschiderea și salvarea fișierelor dialog se va face în aceleași foldere ca și la ultima secvență de lucru similară.

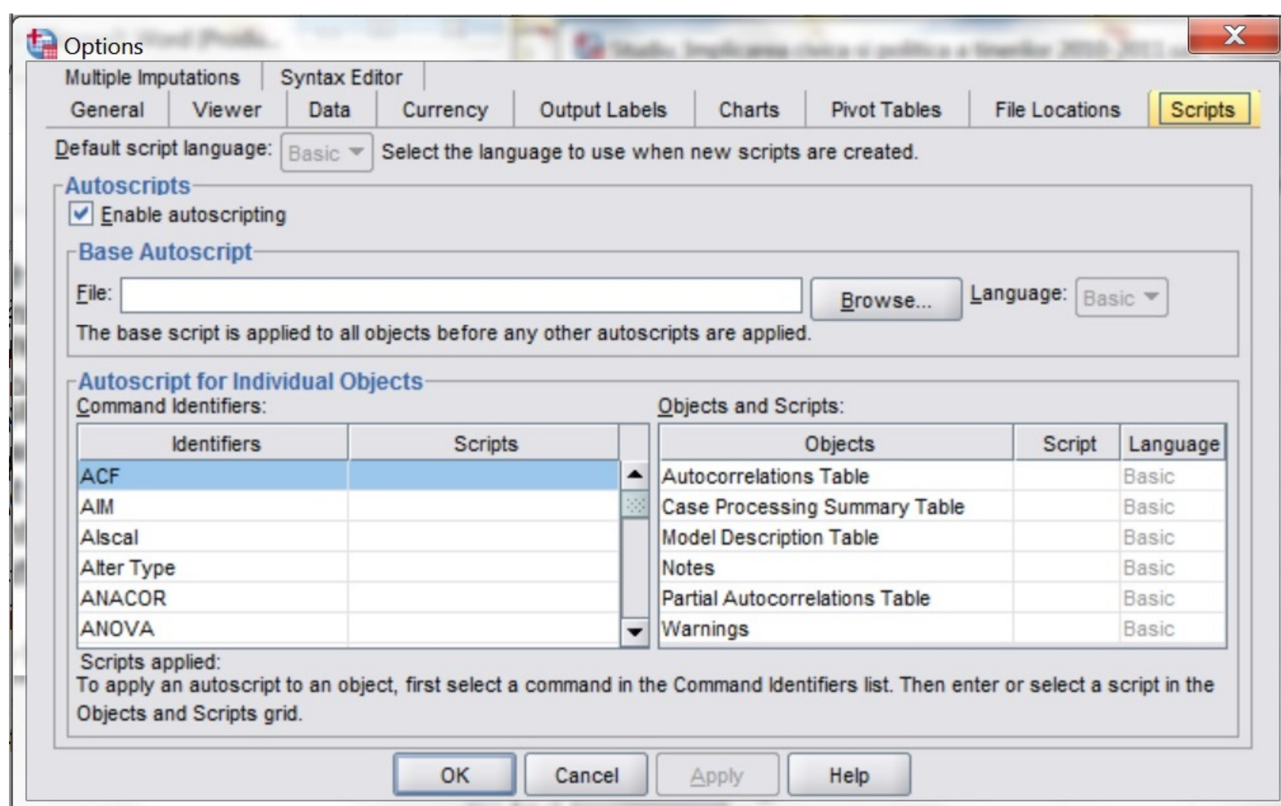
Căsuța de dialog „Session Journal” permte înregistrarea automată a comenzilor executate într-o sesiune („Record syntax in Journal”). Acestea includ comenzi introduse și executate în ferestre de sintaxă și comenzi generate de opțiunile casetei de dialog. Se poate dezactiva și reporni înregistrarea în jurnal, se poate opta pentru adăugarea comenzilor (Append) sau se poate alege înlocuirirea celor vechi („Over write”). Pentru fișierul jurnal se pot selecta numele și locul unde să fie poziționat („Journal File”). Sintaxa cu comanzi SPSS din jurnal se poate copia și salva într-un fișier sintaxă (IBM SPSS Statistics 20 Core System, User’s Guide).

Căsuța de dialog „Temporary folder” specifică locul în care se află fișierele temporare create în timpul unei sesiuni.

Opțiunea „Numbers of Recently Used Files to List” permite alegerea numărului de fișiere care să fie indicate în lista de fișiere utilizate recent.

Comanda „Edit Options Scripts” facilitează setarea utilizării secvențelor automate de comenzi SPSS de tipul „scripts” care au extensia .wwd (IBM SPSS Statistics 20 Core System, User’s Guide).

Graficul 4.15 Comanda Edit Options Scripts în SPSS



În fereastra de dialog „Autoscripts” putem activa secvențe automate de program SPSS („Enable Autoscripting”) care se vor iniția ori de câte ori sunt generate output-uri (prin comenzi SPSS). De exemplu, putem utiliza autoscriptul „ExportTablesToExcelFiles.wwd” (din folderul având calea de acces, pentru SPSS 20, C:\Program Files\IBM\SPSS\Statistics\20\Samples) pentru a exporta tabelele din output-ul SPSS în programul EXCEL. Scriptul automat va fi aplicat înaintea oricăror alte script-uri definite suplimentar (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

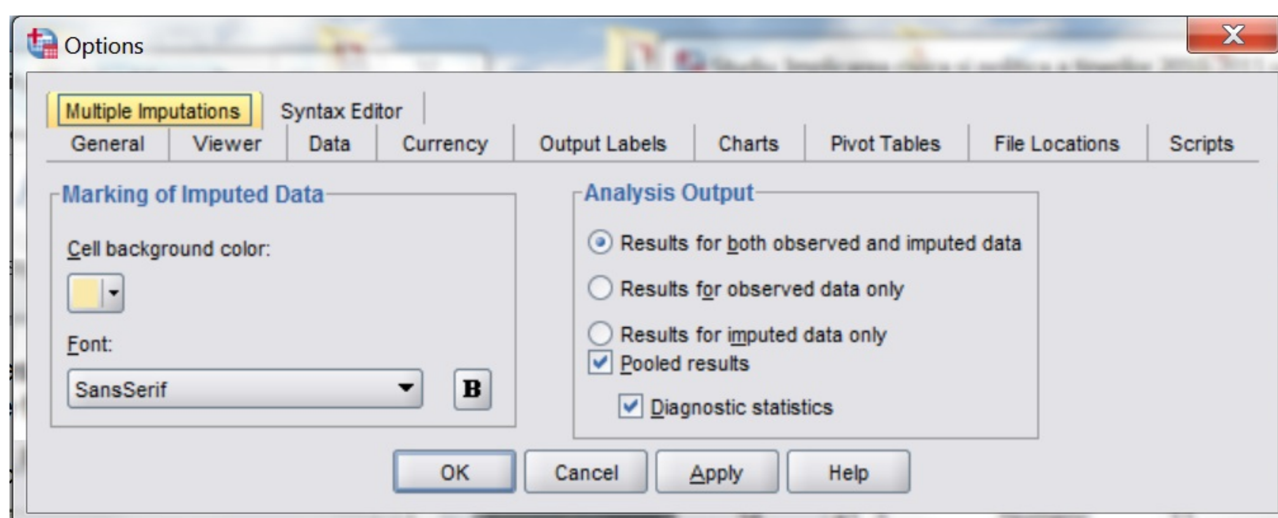
În căsuța „Autoscript for individual objects” putem alege, în prima coloană, o comandă care conduce la un obiect de tip output sau la un script. Selectând comanda respectivă, de ex. „Crosstabs” (din căsuța „identifiers”), vor apărea, în coloana „Objects and Scripts” toate out-puturile și fișierele script care pot fi generate ca succesiuni ale respectivei comenzi.

Am optat ca pentru „Chi Scuare Test table” să alegem script-ul „Swap Rows and Columns Autoscript.wwd” care va schimba între ele liniile și coloanele tabelului.

Pentru „Crostabulation table” (din coloana „Objects”) am optat pentru script-ul „Undo Scientific Notation.wwd” care va anula utilizarea notației științifice pentru numerele din interiorul tabelelor de asociere (pentru SPSS 20, ambele fișiere Script au calea de acces C:\Program Files\IBM\SPSS\Statistics\20\Samples).

Comanda **Edit Options Multiple Imputations** permite setarea modului de operare cu valorile plauzibie generate pentru a înlocui datele lipsă. Acele tipuri de analiză care utilizează mai multe seturi de date care înlocuiesc (prin „imput-are”) informațiile lipsă produc rezultate pentru fiecare set de date „complet”, și, în plus, generează un output unic care estimează ce rezultat s-ar fi obținut dacă setul de date original nu avea valori lipsă. Rezultatele obținute prin sinteza seturilor de date „imput-ate” sunt mai exacte decât cele obținute prin metode de „imput-are” unice. (IBM Knowledge Center, Home, SPSS Statistics 24.0.0)

Graficul 4.16 Comanda Edit Options Multiple Imputations în SPSS



Datele „imput-ate” (care înlocuiesc datele lipsă) pot fi marcate prin culoarea de fond a celulelor (Cell backgrownd Color) și / sau prin alegerea unui anumit font (Font:) respectiv îngroșarea caracterelor (**B**).

Casuța de dialog „Analysis Output” permite, dacă programul SPSS operează cu „imput-ări” multiple ale datelor, să alegem ca în fișierul Output să fie prezentate:

- rezultatele atât pentru datele observate cât și pentru cele „imput-ate”,
- doar output-urile pentru datele observate sau
- numai rezultatele pentru datele „imput-ate”.

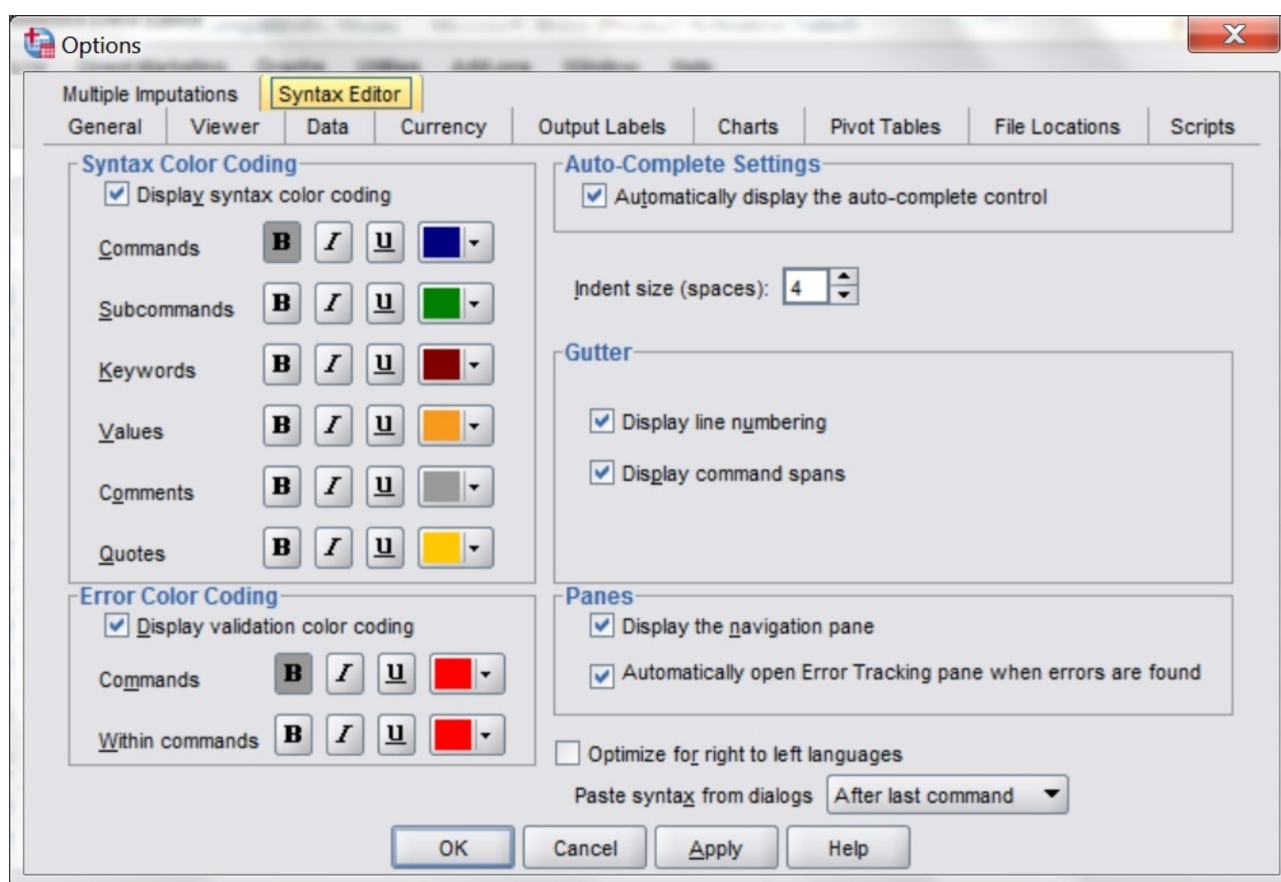
În aceeași căsuță de dialog putem seta afișarea rezultatului sintetic, aproximând situația inexisenței datelor lipsă - Pooled rezults - și diagnoza variabilelor pentru care au fost înlocuite datele lipsă - Diagnostic statistics (IBM SPSS Statistics 20 Core System, User’s Guide).

Comanda „**Edit Options Syntax Editor**” vizează anumite setări care facilitează operarea cu editorul de sintaxă în SPSS.

Căsuța „Syntax Color Codind” se referă la modul de codare al sintaxei de comandă din programul SPSS.

Comenzile, subcomenzile, cuvintele cheie și valorile cuvintelor cheie sunt codate în culori, astfel încât să poată fi identificați ușor termenii necunoscuți. În plus, o serie de erori sintactice comune - cum ar fi ghilimelele neînchise - sunt codate prin culori pentru o identificare mai ușoară (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 4.17 Comanda Edit Options Syntax Editor în SPSS



Căsuța de dialog „Auto-complete Settings” permite afișarea automată sau dezactivarea listei ce dă posibilitatea finalizării automate a comenzilor, subcomenzilor, cuvintelor cheie și valorilor cuvintelor cheie în cadrul fișierelor cu sintaxă de comandă din programul SPSS (IBM SPSS Statistics 20 Core System, User's Guide).

Opțiunea „Gutter” permite setarea implicită pentru afișarea sau ascunderea numerelor liniilor de comandă („Dispaly line numbering”) și a zonei pe care se întinde o linie de comandă („Display commnand span”) în „Gutter-ul” Editorului de sintaxă (regiunea din stânga panoului text rezervată pentru numerele liniilor, pentru marcaje, puncte de întrerupere și intervalul de comandă – vizualizat prin indicatori ai începutului și sfârșitului unei comenzi).

Opțiunea „Panes” permite afișarea Panoului de navigare (Dispaly the navigation pane) respectiv lansarea automată a panoului de urmărire a erorilor: „automatically open Error Tracking pane when errors are found” (IBM SPSS Statistics 20 Core System, User's Guide).

De notat că Panoul de navigare conține o listă a tuturor comenzilor recunoscute din fereastra de sintaxă, afișate în ordinea în care acestea apar. Dacă se dă click pe o comandă din panoul de navigare, cursorul va fi poziționat la începutul comenzii.

Selectarea comenzii „Optimize for right to left languages" permite ca utilizatorul programului SPSS să lucreze în mod optim atunci când sunt folosite simultan două limbaje pentru cele două ferestre ale Panoului sintaxă (care putea fi deschis prin comanzile „File Open Syntax" sau „File New Syntax").

Opțiunea „Pasting Syntax from Dialog Boxes" oferă cel mai simplu mod de a construi un fișier sintaxă de comandă prin selectarea secvențelor de comenzi SPSS din căsuțele de dialog și inserarea acestora într-o fereastră de sintaxă unică. Prin inserarea sintaxei la fiecare pas al unei analize de lungă durată („At cursor or selection"), se poate crea un fișier de lucru care să permită repetarea succesiunii comenzilor SPSS din căsuțele de dialog utilizate. Se poate alege, de asemenea, opțiunea „After last command", care permite lipirea sintaxei în poziția în care se află cursorului din fișierul sintaxă sau supra-scrierea sintaxei selectate (IBM SPSS Statistics 20 Core System, User's Guide).

În capitolul următor ne vom ocupa de facilitățile oferite de meniul „Wiew" în SPSS și PSPP încecând să relevăm aspectele comune proprii celor două programe cât și cele specifice.

5. MENIUL VIEW ÎN SPSS ȘI PSPP

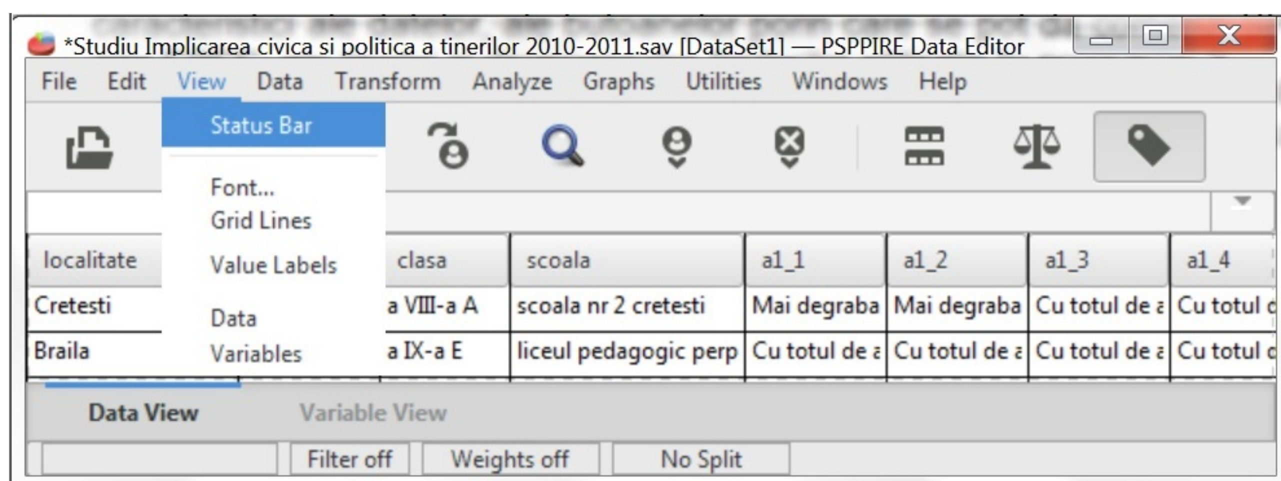
5.1 Opțiunile VIEW Status Bar, Font, Grid lines, Value labels, Data, Variables, în PSPP

Meniul **VIEW** oferă o serie de facilități legate de vizualizarea anumitor caracteristici ale datelor, ale butoanelor prin care se pot da comenzile SPSS sau PSPP respectiv ale tipului și dimensiunilor fonturilor utilizate în programe.

Începem prezentarea cu programul PSPP căci aici există mai puține opțiuni (pe care le vom putea mai ușor compara cu cele oferite de SPSS).

Comanda **View Status Bar,** în PSPP, are rolul de a face vizibile sau de a ascunde informații despre existența unor selectii aplicate asupra unităților statistice existente în baza de date („Filter Off"), a unor ponderări ale valorilor („Weights off") sau a împărțirii bazei de date în mai multe părți („No split") corespunzând categoriilor unei variabile (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1)

Graficul 5.1 Meniul View în PSPP



Selectarea opțiunii „Status Bar" ne va permite să cunoaștem dacă e analizată statistic întreaga bază de date, situație în care apare, în josul ferestrei, mesajul „Filter off" sau ne va da posibilitatea să cunoaștem variabila în funcție de care are loc selecția (pentru valorile unitare ale respectivei variabile entitățile statistice care corespund acestora vor fi selectate în analiză iar pentru valorile 0, vor fi excluse). Tot prin opțiunea „Status Bar" putem să aflăm dacă datele statistice vor fi analizate exact în modul în care au fost introduse în baza de date („Weights off") sau există o variabilă cu ajutorul căreia datele sunt ponderate (situație în care în bara de jos a ferestrei apare numele variabilei de ponderare). Prin aceeași comandă „View Status Bar" putem alege să fim informați dacă în baza de date este activă sau nu comanda „Data Splite File" care are rolul de a împărți baza de date în mai multe componente, corespunzătoare categoriilor unei variabile categoriale. Dacă respectiva comanda este activă, va fi afișată variabila în funcție de care s-a făcut segmentarea bazei de date; în caz contrar, se va afișa „No split".

Comanda **Wiew Fonts** permite alegerea tipului de literă utilizat în vizualizarea barei de date: „Data view" respectiv „Variable view" (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

Opțiunea **Grid Lines** oferă posibilitatea vizualizării sau ascunderii liniilor și coloanelor din baza de date (în cele 2 formate: „Data view" / „Variable view").

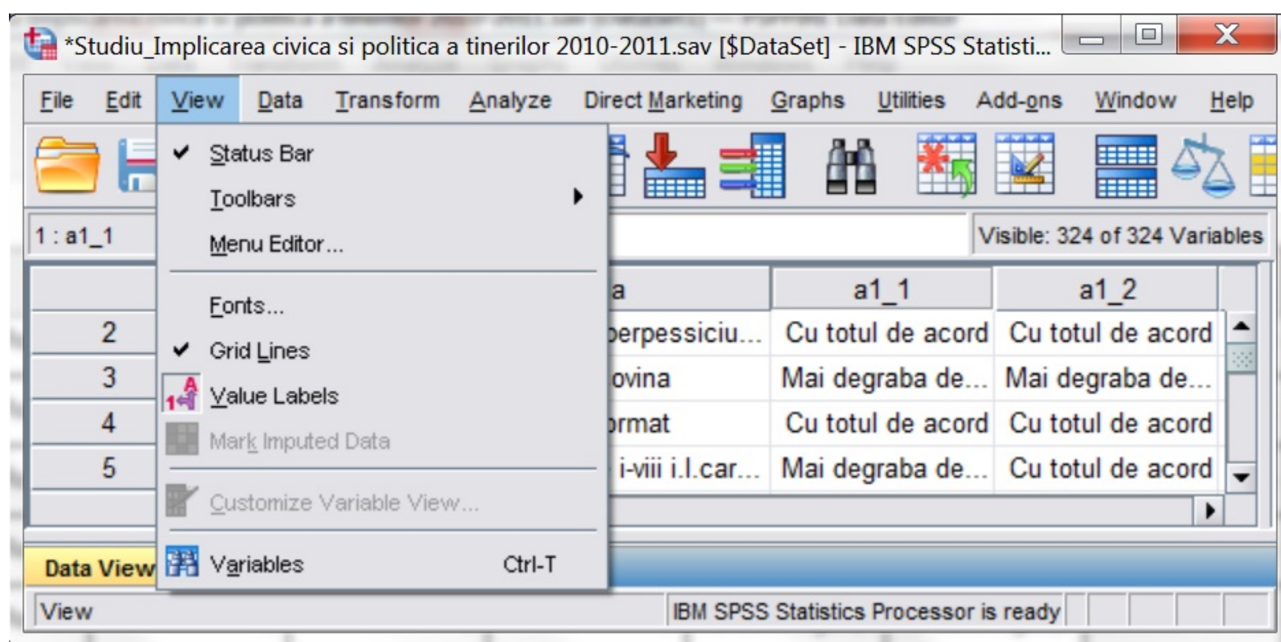
Selectarea sau deselectarea opținii **Value labels** va determina ca în baza de date („Data view" și „Variable view") să fie vizualizate etichetele valorilor variabilelor respectiv valorile efectiv introduse în baza de date.

Selectând opțiunea „Data" respectiv „Label" se poate opta, în primul caz, pentru formatul „Data view" al bazei de date (vizualizarea datelor introduse) iar în al doilea, pentru vizualizarea proprietăților variabilelor (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

5.2 Opțiunile VIEW, Status bar, Toolbars, în SPSS

În SPSS, meniul **View** are ca primă opțiune „**Status Bar**", care activează / dezactivează bara de jos a ferestrei de dialog, sursa de informații despre starea procesorului din SPSS („Processor is ready" dacă a fost finalizată ultima comandă SPSS sau informații despre secvența actuală de lucru).

Graficul 5.2 Meniul View în SPSS



Comanda „**View**", „**Toolbars**", permite selectarea icon-ilor vizualizați în partea de sus a ferestrei de dialog, sub bara cu meniurile SPSS. Iconii sunt sub-comanzi SPSS și pot fi grupați în bare de instrumente (Toolbars) pre-definite sau nou-construite (acestora li se dau nume și pot fi selectate / deselectate).

Urmând comanda „View" „Toolbars", deschidem o fereastră de dialog în care apar barele de instrumente disponibile în Editorul de date SPSS („Data Editor", „Millea" și „Vlad"; de notat că doar prima este activă, cea marcată cu o săgeată neagră). Acest lucru înseamnă că bara de instrumente (sub-comenzi din SPSS) care va apărea în fereastra generală a programului SPSS va fi „Data Editor", conținând sub-comenzile care au fost selectate prin intermediul instrucțiunilor „View", „Toolbars" „Customize" „Data Editor", „Edit". Menționăm că barele de instrumente active (de tipul „Data Editor") puteau fi afișate nu doar în fereastra principală SPSS, ci și în fișierele de tip „Syntax" sau în cele de tip „Output" (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

În graficul 5.3, în partea de jos a ferestrei de dialog „Window” (cu variantele „Data editor”, „Script” și „Output”) putem alege dacă poziționarea mouse-ului deasupra simbolului comenzii SPSS va genera un text explicativ (Show ToolTyps). În aceeași zonă a ferestrei putem alege dacă iconii folosiți pentru comenzile SPSS să fie de dimensiuni mari sau mici („Large buttons”).

Graficul 5.3 Meniul View, opțiunea Toolbars, Customize în SPSS



Dacă în situația din graficul 5.3 dăm comanda „Edit”, vom putea seta care iconi să apară în bara de instrucțiuni „Data Editor”, pe care am selectat-o cu click în coloana „Name” (indiferent dacă aceasta era marcată sau nu în coloana „Show”).

Graficul 5.4 Meniul View, opțiunea Toolbars, Customize, Edit în SPSS



Din ultima fereastră deschisă (dreapta graficului 5.4) putem alege tipul de comenzi (Categories), putem selecta iconii doriți (din Tools) pe care îi vom deplasa în bara de jos - Customizing Toolbar (ce va fi situată în Data Editor). Opțiunea „Reset Toolbars” (baza ferestrei din dreapta graficului 5.4) permite resetarea barei de instrucțiuni, adică readucerea iconilor la forma inițială, din momentul instalării SPSS (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Pentru cele două bare de comenzi definite („Millea” și „Vlad”), opțiunea „Edit” (precedată de „View” „Toolbars”, „Customize”) deschide o fereastră de dialog în care:

- Avem opțiunea de ștergere a tuturor iconilor („Clear Toolbars”) dacă după „View” „Toolbars” selectăm oricare combinație de cel mult 2 din 3 bare de comenzi
- Avem opțiunea de resetare a iconilor („Reset”) și înlocure cu setarea inițială dacă după „View” „Toolbars” am selectat toate cele trei bare de instrucțiuni: Data Editor, „Millea” și „Vlad”)

Opțiunea „Toolbars Proprieties” (dreapta jos în graficul 5.4) permite să alegem tipurile de ferestre principale în care să fie active barele cu iconi reprezentând instrucționi / comenzi SPSS (implicite sau definite de noi): „Data Editor” (gestionează, datele în SPSS), „Viewer” (gestionează datele de ieșire) și „Syntax” (contolează fișierele conținând linii de program SPSS).

În aceeași fereastră de dialog, opțiunea „New Tools” permite selectarea unei instrucțiuni (comenzi) din programul SPSS (IBM, SPSS, Statistics, 20) și accesarea acesteia prin intermediul unui icon (a cărui imagine trebuie definită).

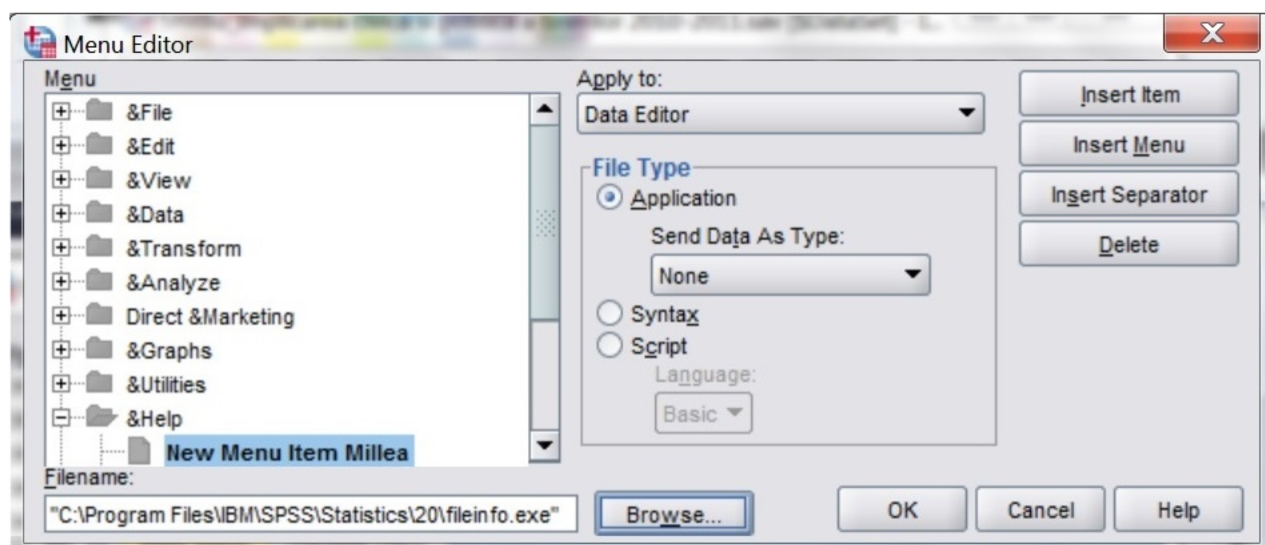
Folosirea comenzilor „View”, „Toolbars”, „Customize” permite să alegem, pentru fiecare fereastră principală („Data Editor” „Viewer” și „Syntax”), care dintre barele de instrumente pre-existente sau definite de noi să fie active.

Succesiunea de comenzi „View”, „Toolbars”, „Customize” „New” permite setarea unei noi bare de instrumente (comenzi), stabilirea unui nume pentru acesta și setarea tipurilor de ferestre principale („Data Editor” „Viewer” și „Syntax”) în care va fi activă respectiva bara de instrucțiuni / comenzi SPSS.

5.3 Opțiunile VIEW, Menu Editor, Fonts, Mark imputed data și Customize Variable View în SPSS

Opțiunea **View, Menu Editor** se poate folosi pentru a da o formă particulară meniurilor principale din SPSS. Se pot adăuga elemente de meniu care execută secvențe „Script” specifice ori rulează fișiere sintaxă de comandă. Noi secvențe de meniu pot lansa programe și trimite date altor aplicații (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 5.5 Meniul View, opțiunea Menu Editor, în SPSS



În partrea dreaptă a ferestrei de dialog obținută prin comanda „View” „Menu editor” se pote observa că sunt prezente toate meniurile din „Data Editor” (mediul de gestionare a datelor în programul SPSS). Pe lîngă meniurile principale sunt prezente o serie de sub-meniuri (accesibile dând click pe patratele cu cruce în mijloc) sau chiar sub-sub-meniuri (în cazul meniului Analyze).

Fereastra „Menu Editor” ne permite să introducem un nou meniu, dacă selectăm unul deja existent și dăm comanda „Insert Menu” din stânga ferestrei (noul meniu va apărea deasupra celui selectat și i se poate atribui un nume; „Vlad” în cazul nostru particular).

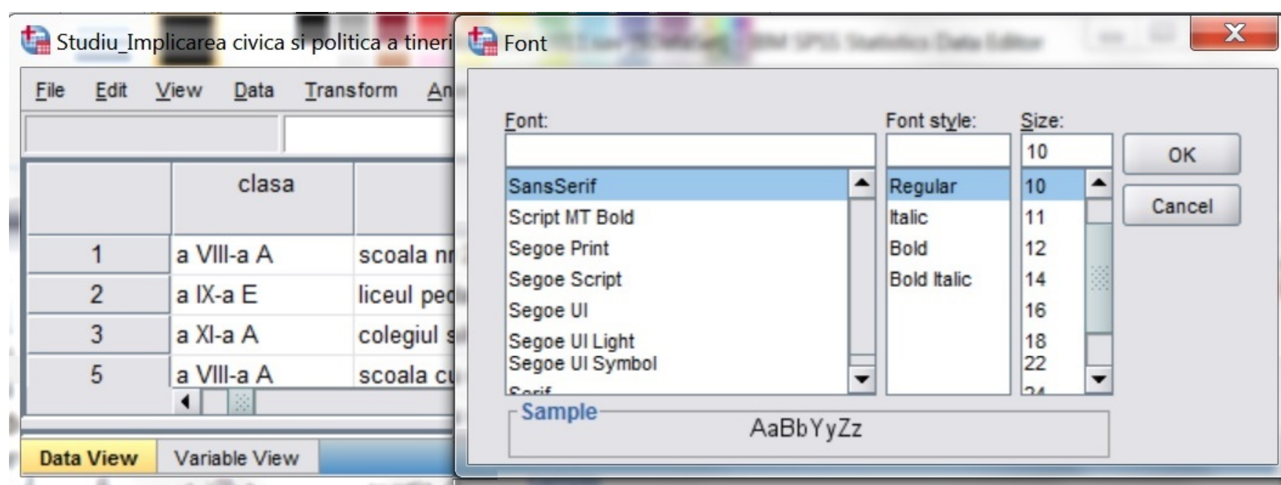
Pentru ca meniul să poată fi validat este nevoie să se introducă cel puțin un sub-meniu (de exemplu, selctăm „End of Vlad Menu” care a fost definit automat) și dăm comanda „Insert Item”. Sub-meniul (numit, de exemplu, „Vlad Menu”) va apărea deasupra folderului anterior selectat și va fi nevoie să indicăm secvența de program SPSS care îi corespunde prin intermediul căsuței de dialog „Filname” - din partea de jos a ferestrei prezentată în graficul 5.5 (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

E important de notat că inserarea de noi meniuri sau sub-meniuri se poate face în „Data Edtor” (care gestionează datele din SPSS), în fișierele de tip Output („Viewer”) sau în editorul sintaxei SPSS („Syntax”).

În graficul 5.5 am selectat primul sub-meniu al meniului „&Help” („To&pics”) și am dat comanda „Insert Item”. Apoi am dat numele noului sub-meniu „New Menu Item Millea” și am indicat succesiunea de pași către secvența de program SPSS care este utilizată de acesta: „C:\Program Files\IBM\SPSS\Statistics\20\fileinfo.exe” (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Succesiunea de comenzi „**View**” „**Fonts**” permite stabilirea tipului de font și dimensiunii acestuia pentru baza de date din SPSS atât în prezentatea „Data View” cât și în prezentarea „Variable View”.

Graficul 5.6 Meniul View, opțiunea Fonts, în SPSS

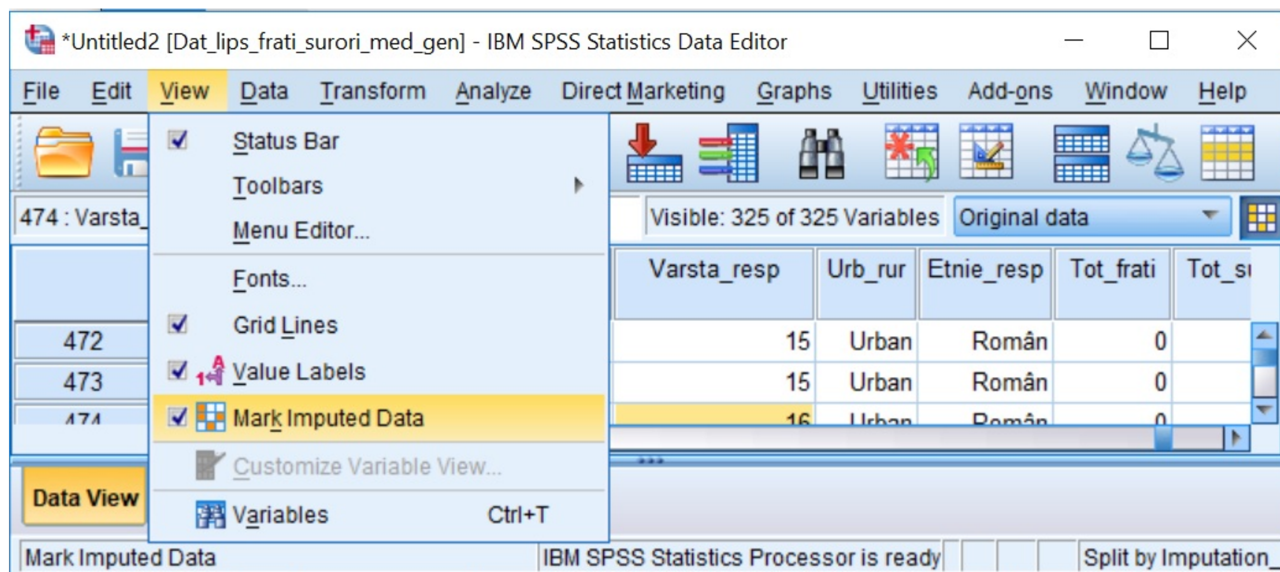


Opțiunea „**View**”, „**Gride Lines**”, în SPSS, acționează identic ca și în programul PSPP, stabilind prezența sau absența liniilor și coloanelor în Editorul de date (varianta „Data view” sau „Variable view”).

Comanda „View” „**Value labels**” stabilește dacă în editorul de date SPSS să fie vizualizate valorile din baza de date sau etichetele acestora (Analysing data using SPSS).

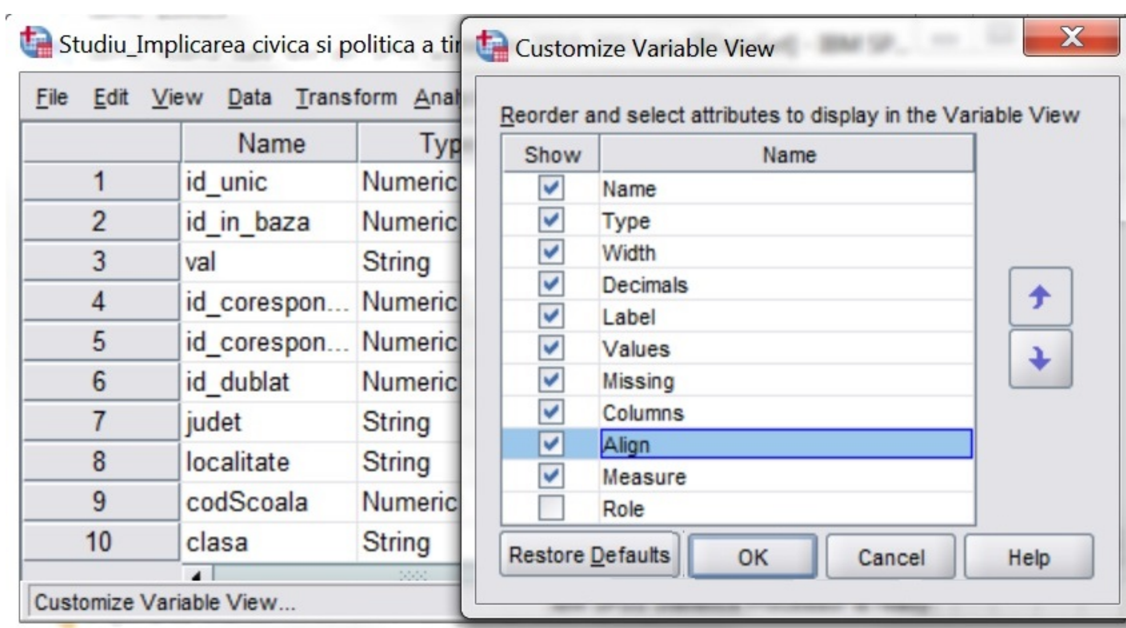
Opțiunea „**Mark imputed data**” din meniul „View” permite marcarea cu culoarea maro a datelor „Imput-ate”, care le-au înlocuit pe cele lipsă în noile baze de date generate de SPSS (de notat că pentru un set de date lipsă pot fi generate mai multe seturi de date „Imput-ate”, ceea ce duce la multiplicarea dmensiunii bazei de date nou-generată)

Graficul 5.7 Meniul View opțiunea Mark imputed data în SPSS



Customize Variable View, din meniul View, ne permite să selectăm proprietățile variabilelor care pot fi setate din fereastra „Variable View” precum și succesiunea acestora (vizualizate în aceeași fereastră).

Graficul 5.8 Meniul View opțiunea Customize Variable View în SPSS



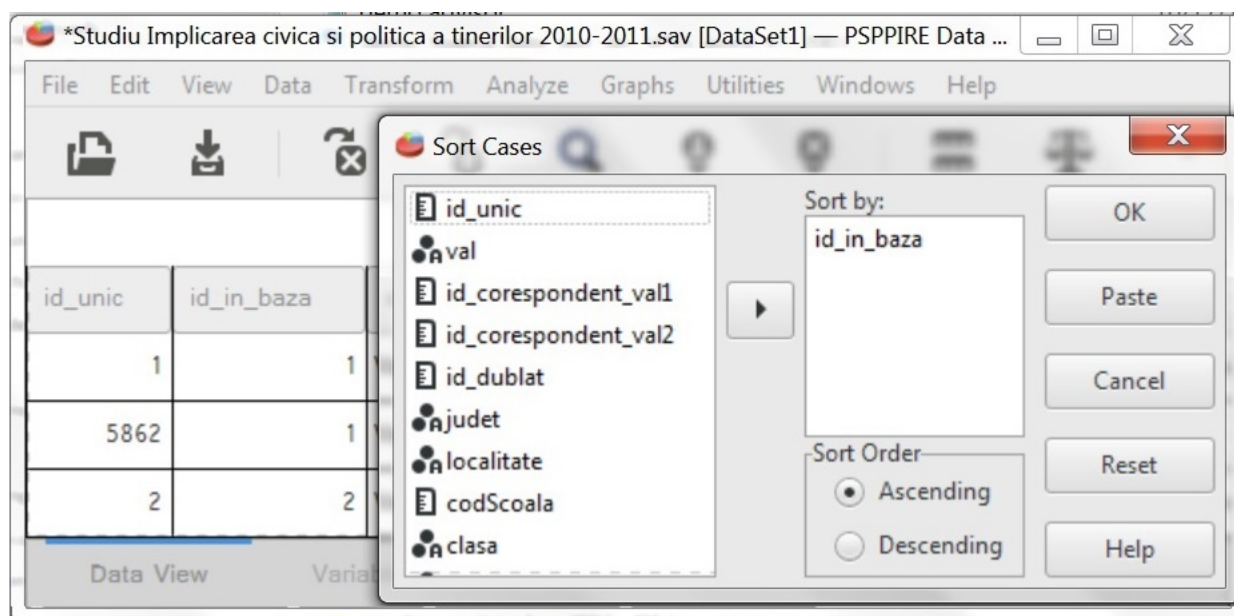
Comanda „View” „Data” determină ca în Data Editor să fie prezentate datele iar comanda „View” „Variables” determină vizualizarea variabilelor (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

6. MENIUL DATA ÎN SPSS ȘI PSPP

6.1 Opțiunea DATA, Sort Cases, Transpoze, Data Agregate în PSPP

Cu ajutorul meniului „**Data**”, opțiunea „**Sort cases**” putem ordona baza de date în funcție de valorile crescătoare ale uneia sau mai multor variabile (se ordonează în funcție de valorile primei variabile apoi, în interiorul cateoriilor acesteia, se ordonează în funcție de a doua variabilă; apoi se ordonează crecător după a treia variabilă în categoriile combinate ale primelor două și așa mai departe, în funcție de numărul total de variabile folosite la sortare).

Graficul 6.1 Meniul Data, Sort Cases în PSPP



Opțiunea **Data Transpoze** permite înlocuirea coloanelor din baza de date cu liniile (cazurile se transformă în variabile și variabilele în cazuri).

Graficul 6.2 Meniul Data Transpoze în PSPP

*[DataSet2] — PSPPIRE Data Editor

Case	Nume	Dan	Maria	Victor	Alexandru
1	Profesor1	4	3	1	3
2	Profesor2	1	4	3	2
3	Profesor3	3	2	4	1
4	Profesor4	2	1	2	4

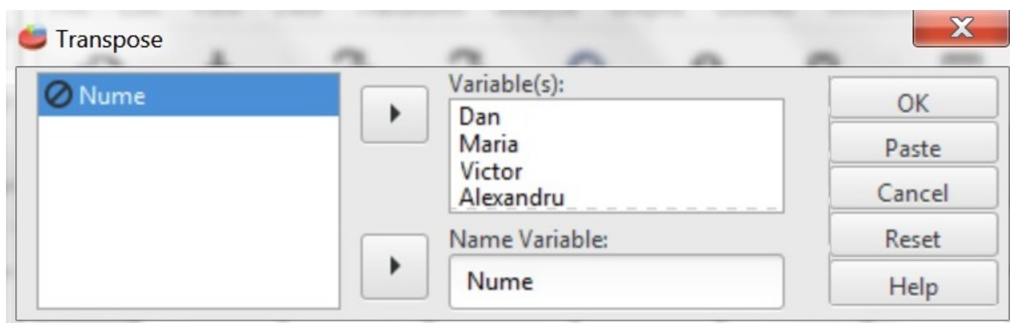
*[DataSet2] — PSPPIRE Data Editor (4 cases × 1 variable)

Case	CASE_LBL	Profesor1	Profesor2	Profesor3	Profesor4
1	Dan	4	1	3	2
2	Maria	3	4	2	1
3	Victor	1	3	4	2
4	Alexandr	3	2	1	4

În căsuța de dialog 6.3 observăm faptul că variabilele cu numele Dan, Maria, Victor și Alexandru s-au transformat îm cazurile cu aceleași nume (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1.)

În căsuța de dialog din graficul 6.3 se poate observa că variabilele numerice din baza de date aflată în stânga graficului 6.2 (Dan, Maria, Victor și Alexandru) au fost mutate în zona „Variable(s)" pe când variabila string „Nume" a fost deplasată în căsuța intulată „Name Variable".

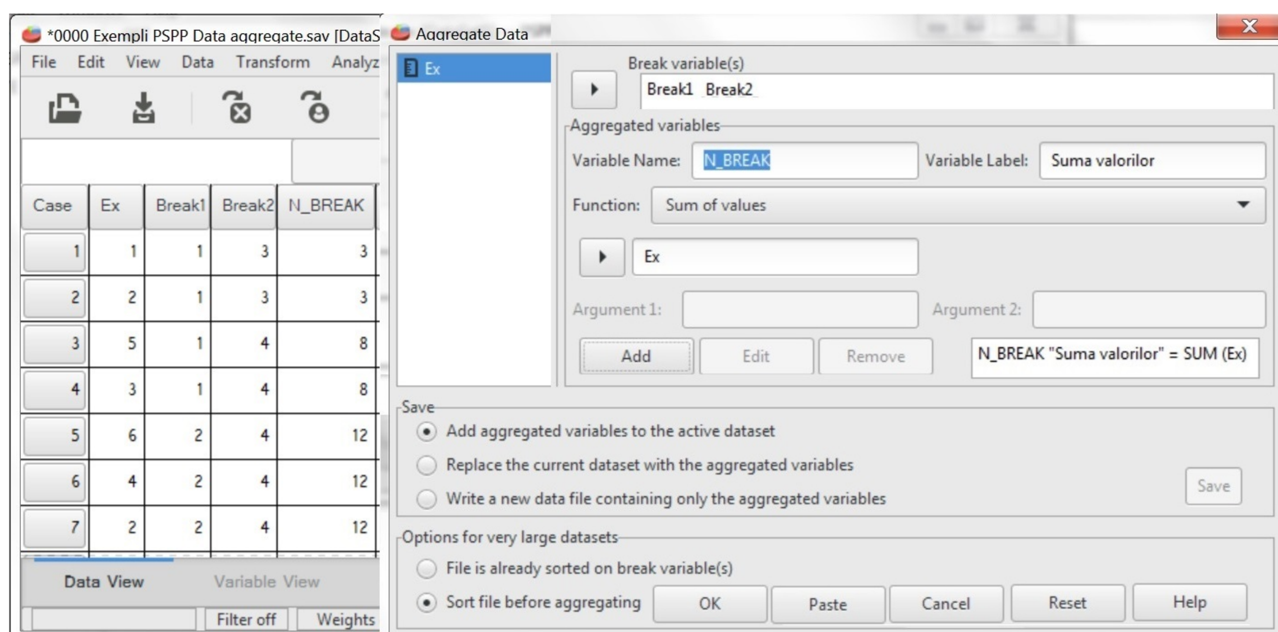
Graficul 6.3 Căsuța de dialg a comanzii Data Transpoze în PSPP



Variabilele numerice Dan, Maria, Victor și Alexandru se regăsesc sub formă de cazuri (linii) în baza de date construită prin comanda „Transpoze" pe când valorile variabilei „Nume" au fost folosite pentru a denumi noile variabile din baza de date aflată în partea dreptă a graficului 6.2 (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

Meniul „**Data**" „**Agregate**", în PSPP, facilitează generarea de valori unice pentru grupuri de indivizi cuprinse într-un set de date, grupuri care sunt delimitate cu ajutorul uneia sau mai multor variabile de tip „Break", care iau aceleași valori pentru grupurile pe care le delimitează (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

Graficul 6.4 Meniul Data Agregate în PSPP



În PSPP există o multitudine de proceduri matematice de agregare, incluzând „Sume of Values" prin care fiecărui set de date pentru care „Break variables" iau aceleași valori i se va atribui valoarea egală cu suma grupului (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

În dreapta graficului 6.4 e prezentată baza de date conținând valorile variabilei denumită „Ex" care, prin intermediul vatiabielor „Break1" și „Break2" este împărțită în în 3 grupuri:

- pentru grupul „Break1"= 1 și „Break2" =3, valoarea pt N_BREAK e 1+2=3
- pentru grupul „Break1"= 1 și „Break2" =4, valoarea pt N_BREAK e 5+3=8
- pentru grupul „Break1"= 2 și „Break2" =4, valoarea pt N_BREAK e 6+4+2=12

Notăm că varabila de agregare N_BREAK a fost generată după ce au fost date comenzile „Data" „Agregate" iar variabilele „Break" au fost introduse în căsuța de dialog din dreapta-sus a graficului 6.4. Variabila „Ex", ale cărei valori tebuiau agregate, a fost introdusă în căsuța de dialog aflată sub cea care permite alegerea procedurii matematice de agregare („Function") și apoi a fost dată comanda „OK" de la butonul din partea de jos.

Important: căsuțele „Argument 1" și „Argument 2" devin active dacă funcția matematică selectată prin „Function" necesită indicarea unei valori numerice (de ex., procentul valorilor din grup aflate sub o anumită valoare).

Pentru tratarea ulterioară a datelor, se poate opta pentru trei variante diferite (vezi căsuța de dialog „Save"):

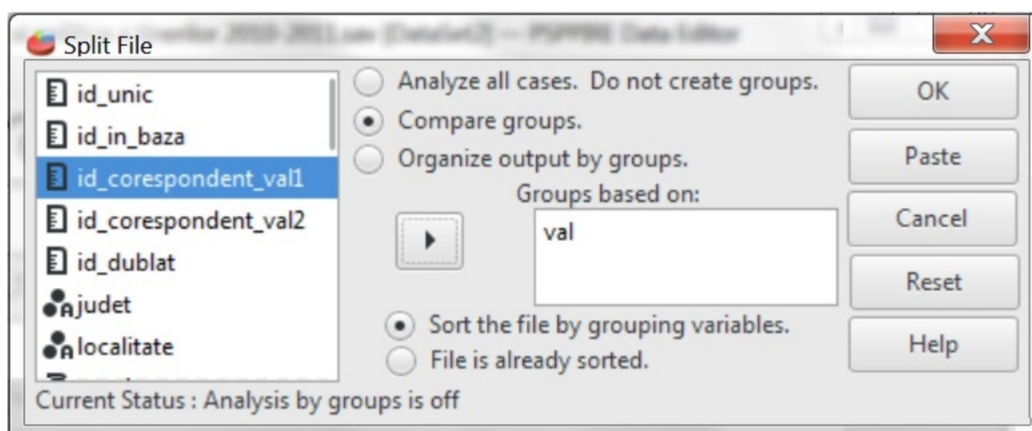
- Adăugarea variabilelor agregate la baza de date activă (extensia „.sav")
- Înlocuirea bazei de date active cu cea formată din date agregate
- Crearea unei noi baze de date conținând datele agregate (buton activ „Save")

Un ultim aspect care trebuie menționat îl constituie opțiunile vizând fișiere de foarte mari dimensiuni (Option for very large datasets) pentru care se poate ordona în prealabil baza de date (Files is already sorted on break variables) sau se poate opta ca programul PSPP să facă respectiva ordonare după ce a fost dată comanda de construcție a variabilei de agregare (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

6.2 Opțiunea DATA, Split file, Select Cases, Weigh Cases în PSPP

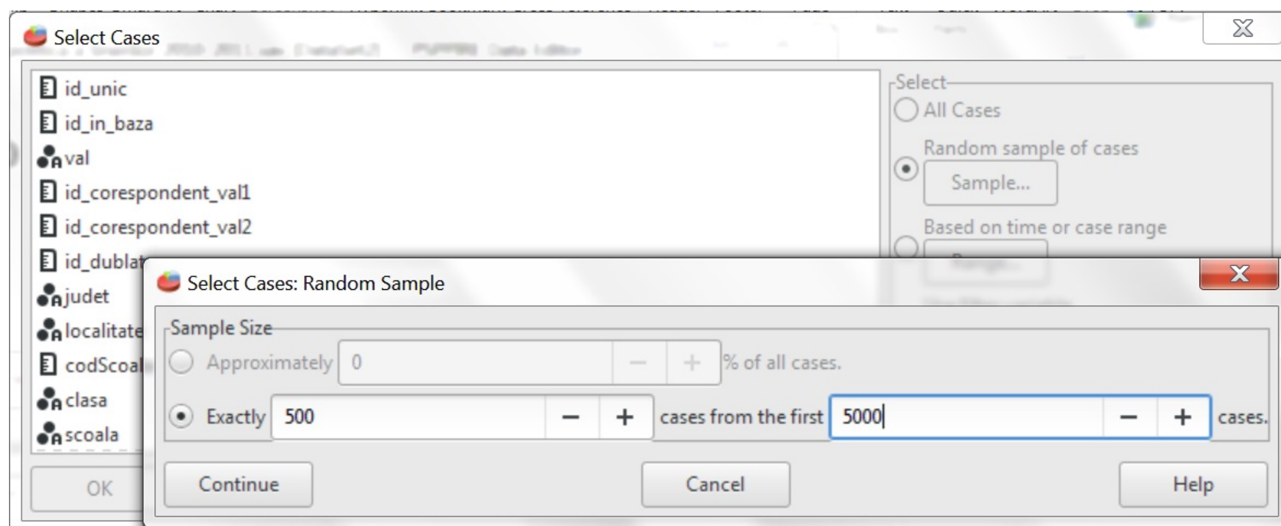
Meniul „**Data**" „**Split File**" permite ca analizele în PSPP să aibă loc în interiorul categoriilor variabilei în funcție de care s-a segmentat baza de date. Tabelele cu date de ieșire pot fi compacte, facilitând comparții (compare groups), sau se pot realiza separat pentru fiecare categorie a variabilei (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

Graficul 6.5 Meniul Data Split file în PSPP



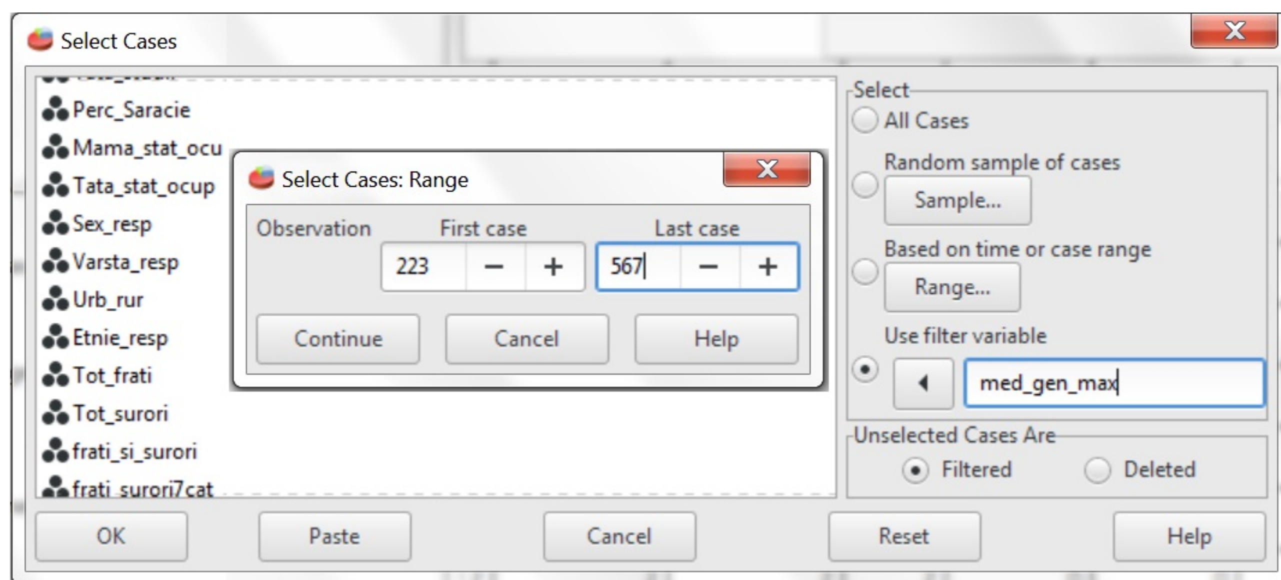
Meniul „**Data**” „**Select Cases**” în PSPP permite selectatrea anumitor cazuri din baza de date sub diferite forme, în funcție de scopul analizei ulterioare.

Graficul 6.6 Meniul Data Select Cases, Random Sample, în PSPP



În graficul 6.6 s-a ales opțiunea „Select” „Random sample of cases” și a fost generată căsuța de dialog din partea de jos a ferestrei inițiale. În Partea de sus a respectivei căsuțe putem alege „Sample Size” „Approximatively” unde vom introduce procentul de cazuri pe care dorim să-l extragă aleator programul PSPP din întreaga bază de date. Dacă, însă, alegem varianta „Exactly”, trebuiei să indicăm numărul de cazuri selectate aleator din primele ”X” cazuri ale bazei de date (se indică „cases from the first ... cases”). Cazurile se pot selecta nealeator cu opțiunea „Based on time or case range” pentru care trebuie indicat primul și ultimul caz (grafic 6.7, stânga).

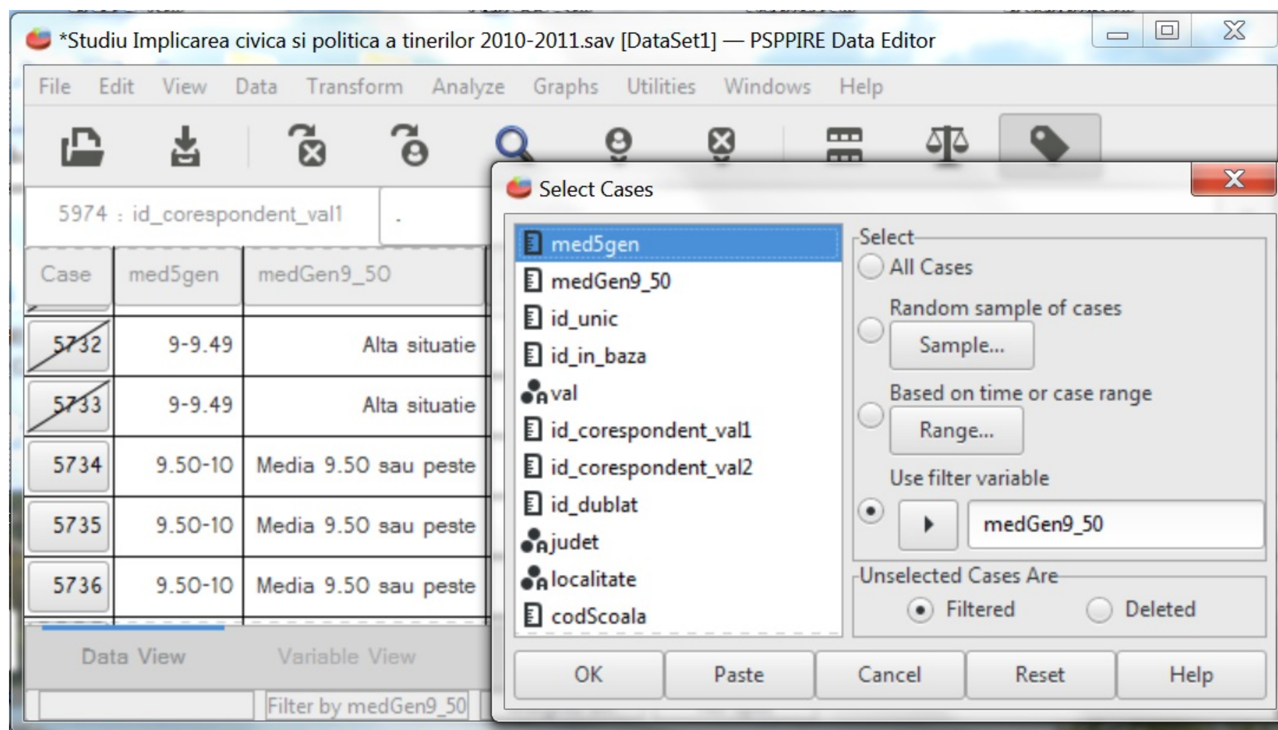
Graficul 6.7 Meniul Data Select Cases, Range în PSPP



Dacă în graficul 6.7 alegem varianta „Use filter variable” va fi nevoie să utilizăm o variabilă care ia valoarea 1 pentru cazurile selectate și 0 în rest (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

Notăm că o astfel de variabilă se poate construi cu ajutorul meniului Transform, opțiunea „Compute" utilizând o expresie logică care să fie adevărată (și să primească valoare numerică 1) pentru cazurile pe care dorim să le selectăm și să ia valoarea 0 în celelalte situații. Atragem atenția că e nevoie să definim datele lipsă pentru variabilele cu care operăm în expresia logică, în caz contrar putând apărea valori improprii ale acesteia (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

Graficul 6.8 Meniul Data Select Cases, Use filter variable în PSPP



În graficul 6.8 se poate vedea că variabila medGen9_50 a fost folosită pentru a selecta cazurile care vor fi supuse analizei în continuare. Respectiva variabilă a fost construită folosind comanda „Transform Compute", utilizând expresia logică „med5gen = 5" (care este adevărată când media generală a elevilor este mai mare sau egală cu 9.50 și falsă în celelalte situații).

În graficul 6.8, în partea stângă, se observă că pentru valorile variabilei „med5gen" egale cu 9-9.49 (valoarea 4 în baza de date), variabila utilizată pentru selecție a luat valoarea „Altă situație" (0, în baza de date) iar căsuțele 5732 și 5733 sunt barate, ceea ce înseamnă că sunt excluse din viitoarele analize. Pentru valorile „med5gen" egale cu 5 (având eticheta 9.50-10), variabila „medGen9_50" ia valoarea egală cu valoarea 1 (în baza de date) iar căsuțele 5734 și 5735 și 5736 nu sunt barate, cee ce înseamnă că doar acele cazuri vor fi analizate statistic după finalizarea comenzii „Select cases".

De notat că după selectarea unui grup de cazuri cu instrucțiunea „Select cases" toate analizele de date din PSPP se vor realiza doar asupra respectivului grup, până în momentul în care selecția va fi revocată prin comanda „Select all cases" (vezi partea de sus a fereastrei de dialog „Select cases" din graficul 6.8). E important de reținut că PSPP oferă informații (în fișierul output) despre filtrarea aplicată însă doar în momentul execuției ei (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

În programul PSPP există posibilitatea de a pondera cazurile existente în funcție de valorile luate de o variabilă definită special în acest scop. Variabila de ponderare va determina programul SPSS să ia în analiză fiecare caz de un număr de ori egal cu valoarea variabilei de ponderare. O astfel de procedură poate fi folosită pentru a însuma mai multe eșantioane aleatoare corespunzând categoriilor unei variabile despre care cunoaștem distribuția reală în populația totală. Dacă volumele sub-eșantioanelor nu sunt în concordanță cu ponderea lor reală în populație, ele pot fi ponderate.

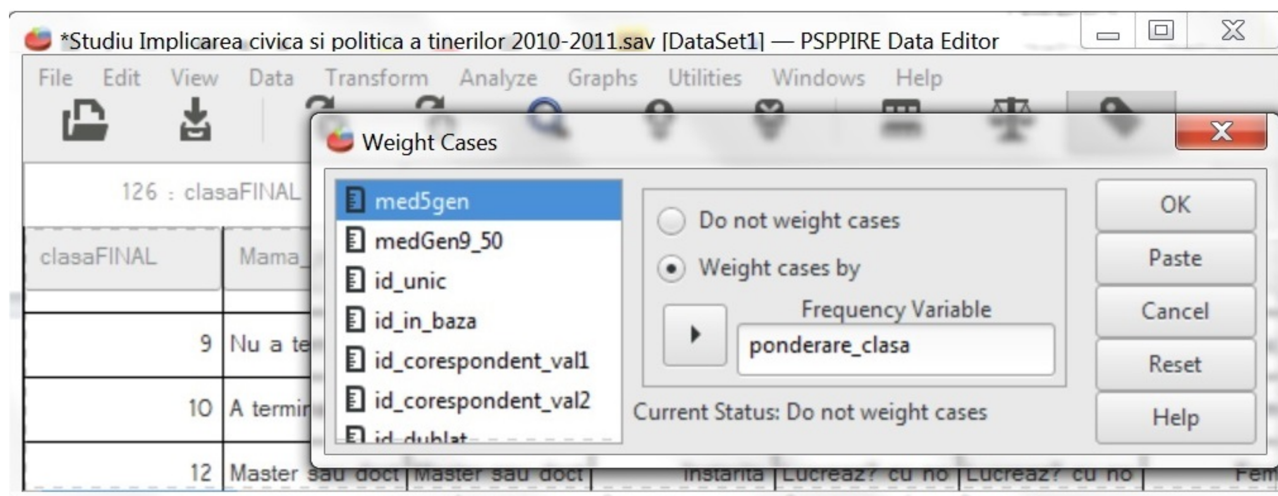
În Graficul 6.9 putem presupune că avem 5 eșantioane probabiliste de elevi corespunzând câte uneia din clasele 8-12, având 938 – 4032 subiecți.

Grafic 6.9 Exemplu Weigh Cases: Clasa in care se află elevul

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	8	1033	13.12	13.13	13.13
	9	4032	51.23	51.23	64.36
	10	919	11.68	11.68	76.04
	11	938	11.92	11.92	87.95
	12	948	12.04	12.05	100.00
NR	99	1	.01	Missing	
	Total	7871	100.0	100.0	

Dacă știm că în realitate fiecare clasă de elevi are aceeași pondere în totalul populației, calculăm valorile de ponderare asfel încât să fie generată, în PSPP, valoarea medie a numărului de cazuri pentru fiecare clasă (1574.2). Pentru clasa a 8-a valoarea va fi 1574.2/1033, pentru a 9-a, 1574.2/4032, etc. În graficul 6.10 e prezentată fereastra prin care se selectează variabila de ponderare („ponderare_clasa”) care va face ca fiecare clasă să aibă aceeași contribuție la numărul total de cazuri din baza de date (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1.)

Graficul 6.10 Meniul Data Weigh Cases în PSPP

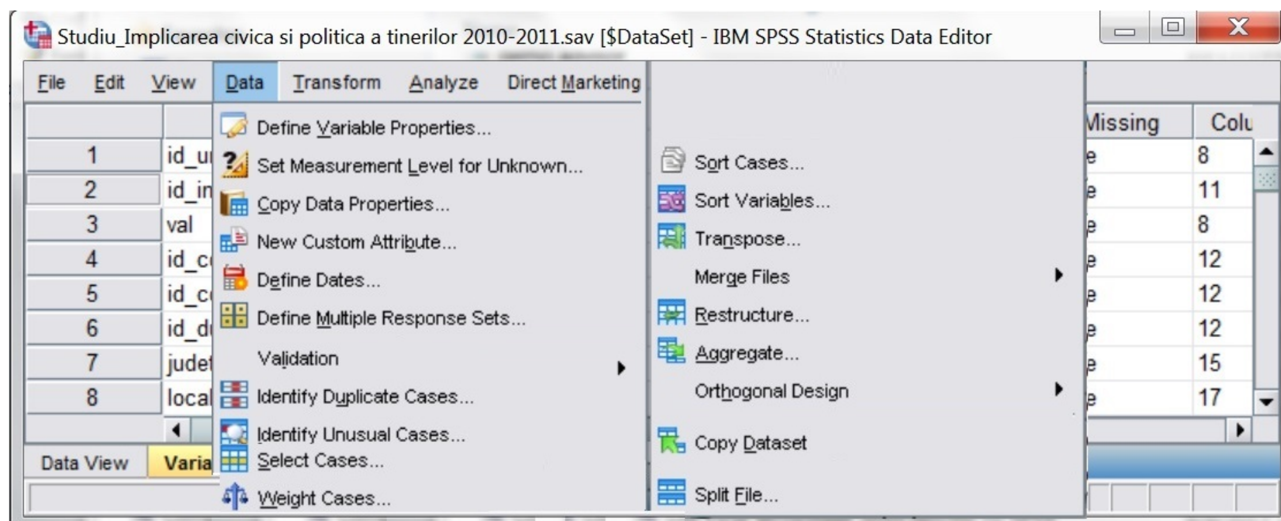


Vom genera variabila de ponderare cu comanda „Compute” construind câte o variabilă care are valoarea de ponderare proprie grupului desemnat de fiecare clasă și 0 în rest. Variabila de ponderare e suma celor 5 variabile (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

În continiare vom descrie facilitățile oferite de meniul DATA din SPSS, care sunt mult mai numeroase decât în PSPP.

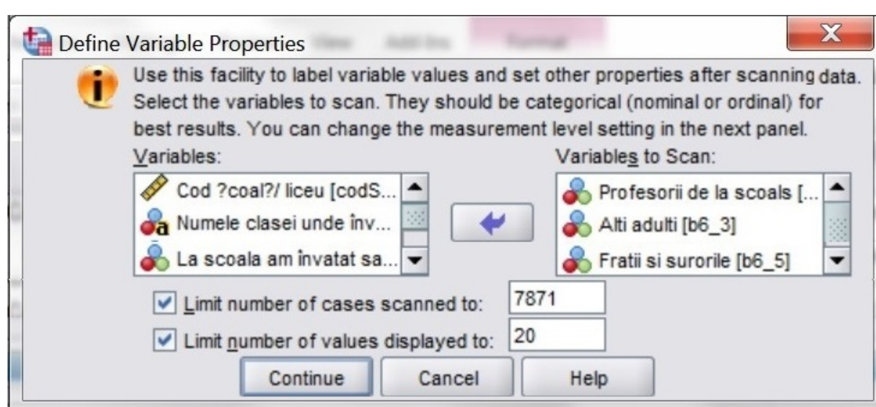
6.3 Opțiunea DATA, Define variable proprieties, Set measurement level for unknown, în SPSS

Graficul 6.11 Meniul Data în SPSS



Fereastra de dialog din graficul 6.12 a fost obținută perin comanda „Data" „**Define variable proprieties**" și e utilizată pentru a eticheta valorile variabilelor și a stabili alte proprietăți după scanarea datelor (în graficul 6.11 vedem că e prima opțiune a meniului DATA în SPSS). Variabilele selectate pentru scanare (nominale sau ordinale) trebuie să aibă un număr limitat de categorii pentru a obține rezultate optime (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

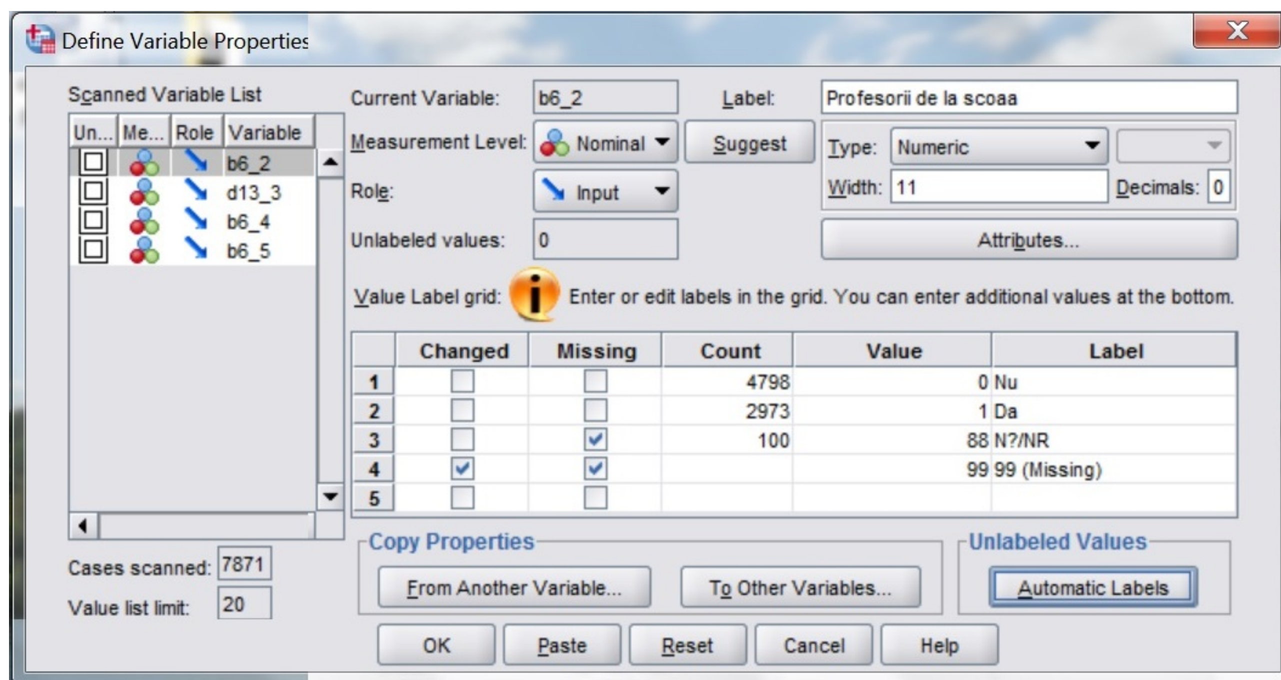
Graficul 6.12 Meniul Data, Define variable proprieties 1 în SPSS



În graficul 6.12 se poate vedea lista variabilelor din baza de date în partea stângă a ferestei de dialog iar în partea dreaptă se regăsesc variabilele care au fost selectate pentru a fi scanate. Tipul variabilei va putea fi modificat în următoarea fereastră (obținută prin comanda „Continue"), după ce în prealabil se stabilește numărul total de cazuri scanate (Limit number of cases scanned to:) și se limitează numărul maxim de valori distincte ale variabilei care sunt luate în analiză – „Limit number of values displayed to:" (Official IBM SPSS Statistics 21 Brief Guide).

În graficul 6.13 este prezentată procedura de definire a proprietăților variabilei cu numele b6.2 și eticheta – care poate fi schimbată – „Profesorii de la școală”. Variabila, definită ca find calitativ nominală, poate fi reevaluată ca ordinală sau cantitativă. Poate fi modificat tipul variabilei (Type) dar comanda nu va avea rezultat dacă o variabilă introdusă în baza de date ca „String” va fi redefinită ca „Numeric” (operația inversă este, însă, posibilă).

Graficul 6.13 Data, Define variable proprieties 2 în SPSS



Variabila poate fi definită (redefinită) ca având următoarele roluri (Role): poate fi dată de intrare („Input”), dată de ieșire („Output”), poate avea dublu rol („Both”) sau poate avea rol de partiționare („Patrition”) și divizare a bazei de date („Split”) în sub-grupuri analizate separat (Official IBM SPSS Statistics 21 Brief Guide).

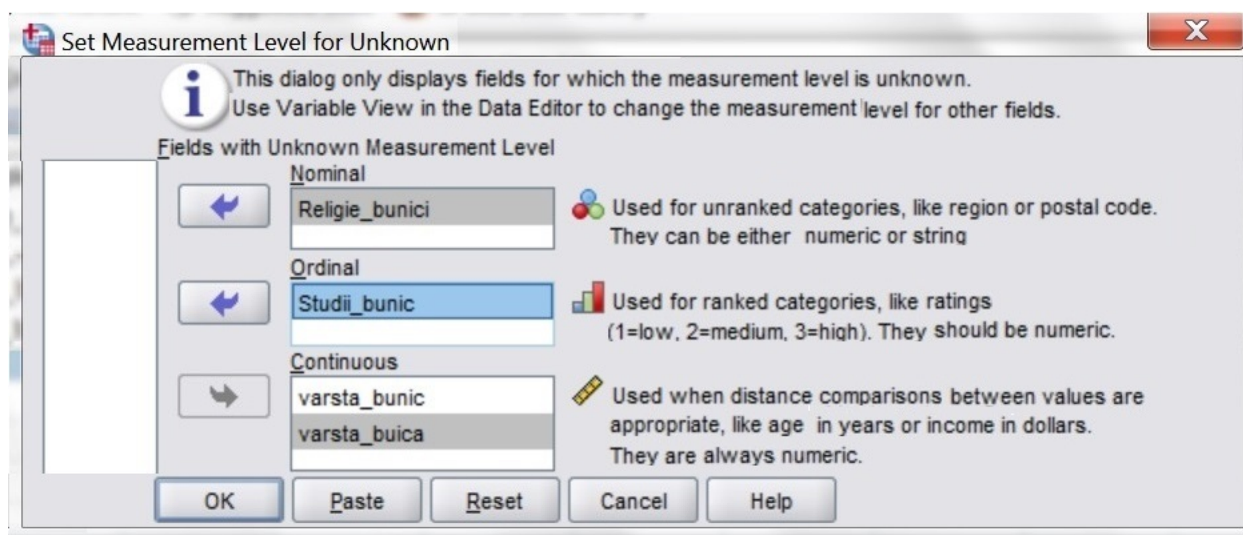
Fereastra semnalează dacă au fost identificate valori ale variabilei cărora nu le-au fost alocate etichete („Unlaibled values”) și dă posibilitatea definirii respectivelor etichete (în coloana „Label”). De asemenea, în fereastră sunt prezentate frecvențele pentru fiecare valoare a variabilei („count”) și se oferă posibilitatea definirii datelor lipsă („Missing”). Coloana „Changed” indică valorile variabilei pentru care s-au făcut shimbări față de starea inițală.

De notat că în partea de sus dreapta a graficului 6.13 se pot schimba urm[toarele aspecte: tipul variabilei, eticheta acesteia (Label:), numărul maxim de caractere (With) și numărul de zecimale (Decimals).

În aceeași zonă a graficului 6.3 există posibilitatea de a defini noi atribute pentru variabila curentă (de exemplu, „valul în care a fost aplicată” sau „întrebări cu răspuns simplu sau multiplu”. Folosind butonul „Attributes” vom putea să introducem numele unei noi caracteristici a variabilelor din baza de date precum și una sau mai multe valori care pot fi luate de variabile în funcție de respectivele caracteristici. Atributul nou-introdus va fi vizibil în varianta „Variable View” a editorului de date SPSS. În final va trebui să indicăm, în coloanele atributelor nou definite, care sunt valorile care corespund pentru fiecare variabilă în parte (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Comanda „**Data**" „**Set measurement level for unknown**" ne permite să identificăm variabilele pentru care n-a fost definit nivelul de măsurare dpdv. sociologic și să-l stabilim. Din stânga graficului vom selecta variabilele fără nivel de măsurare și le introducem în una din zonele centrale, după natura lor (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 6.14 Fereastra Data Set measurement level for unknown SPSS

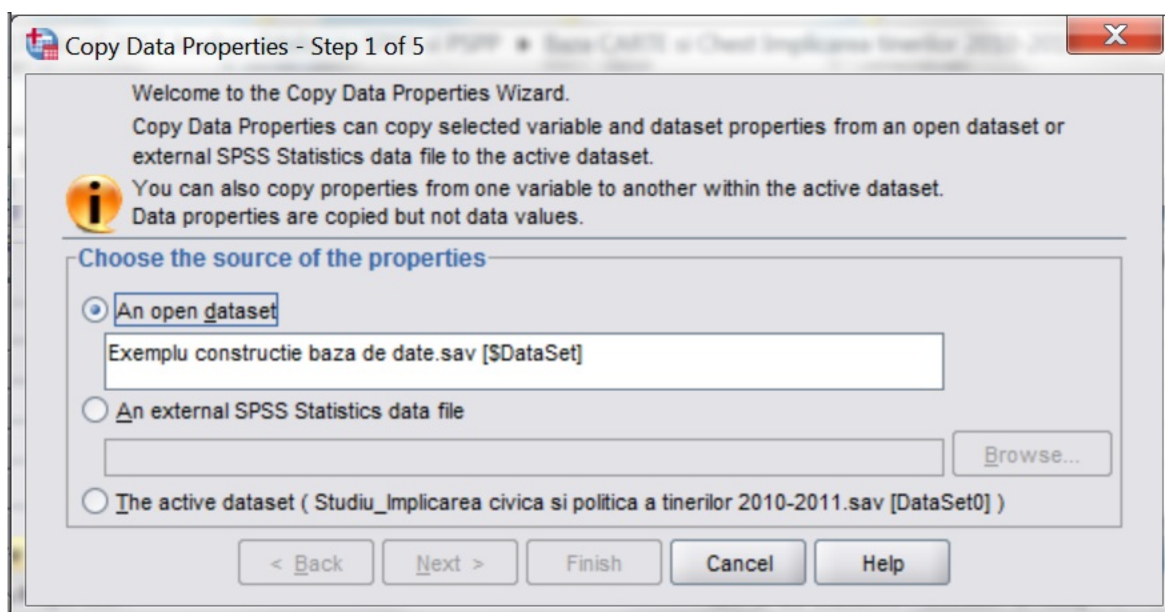


6.4 Variantele DATA, Copy Data Properties New Custom Attribute, Define Dates în SPSS

Cu ajutorul comenzii **Data**, **Copy Data Properties** poate fi copiat modul de definire al unor variabile aflate:

- în altă bază de date deschisă simultan cu cea în lucru (An open dataset)
- într-o bază de date externă de tip SPSS
- în baza de date cu care lucrăm în momentul respectiv (IBM Knowledge Center, Home, SPSS Statistics 24.0)

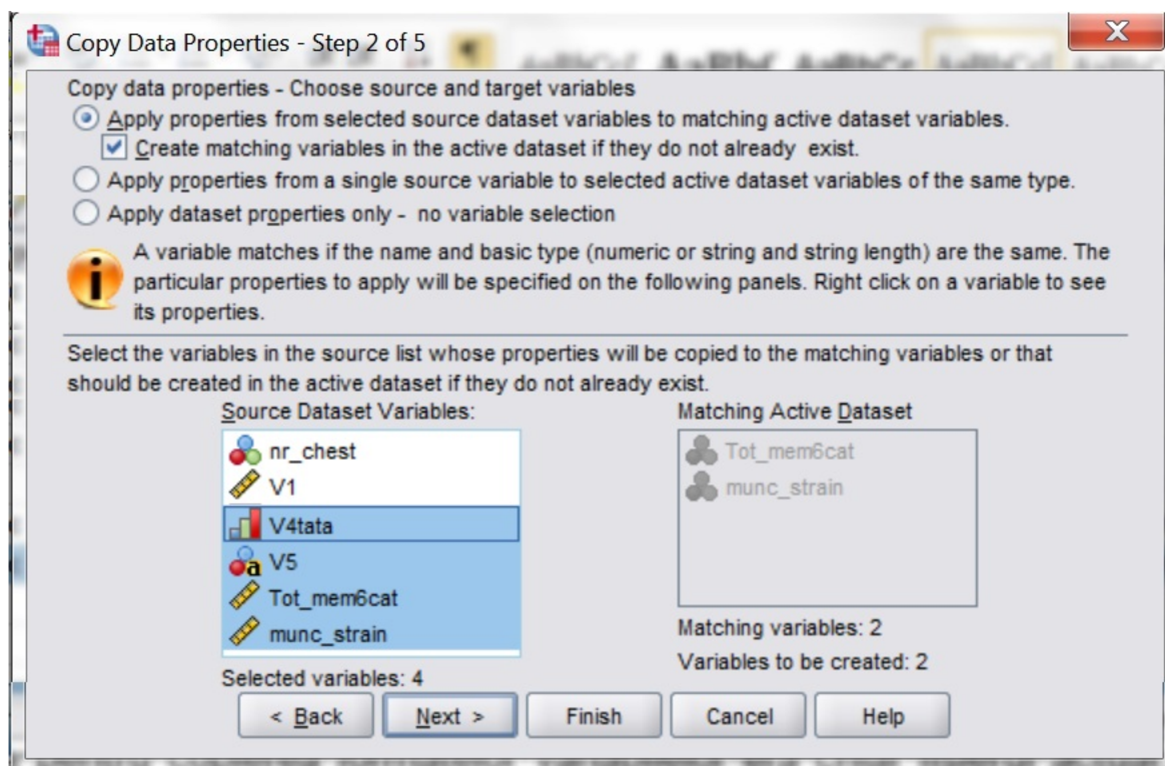
Graficul 6.15 Comanda Copy Data Properties (1) în SPSS



În graficul 6.15 observăm că este activă opțiunea „An open datset" ceea ce înseamnă că se dorește copierea formatului unor variabile (inclusiv a numelor acestora) din alt fișier activ decât cel în care am dat comanda SPSS. Dacă am fi selectat `„An external SPSS Statistic data file" atunci ar fi trebuit să identificăm calea către respectivul fișier folosind căsuța „Browse". Dacă, în schimb, sursa pentru copierea formatelor variabilelor era chiar fișierul actual, atunci trebuia selectată varianta „The active dataset" (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Dacă alegem să copiem formatul variabilelor într-un fișier extern, vom selecta acel fișier tip bază de date vizat și apoi vom da comanda „Next".

Graficul 6.16 Comanda Copy Data Properties (2) în SPSS



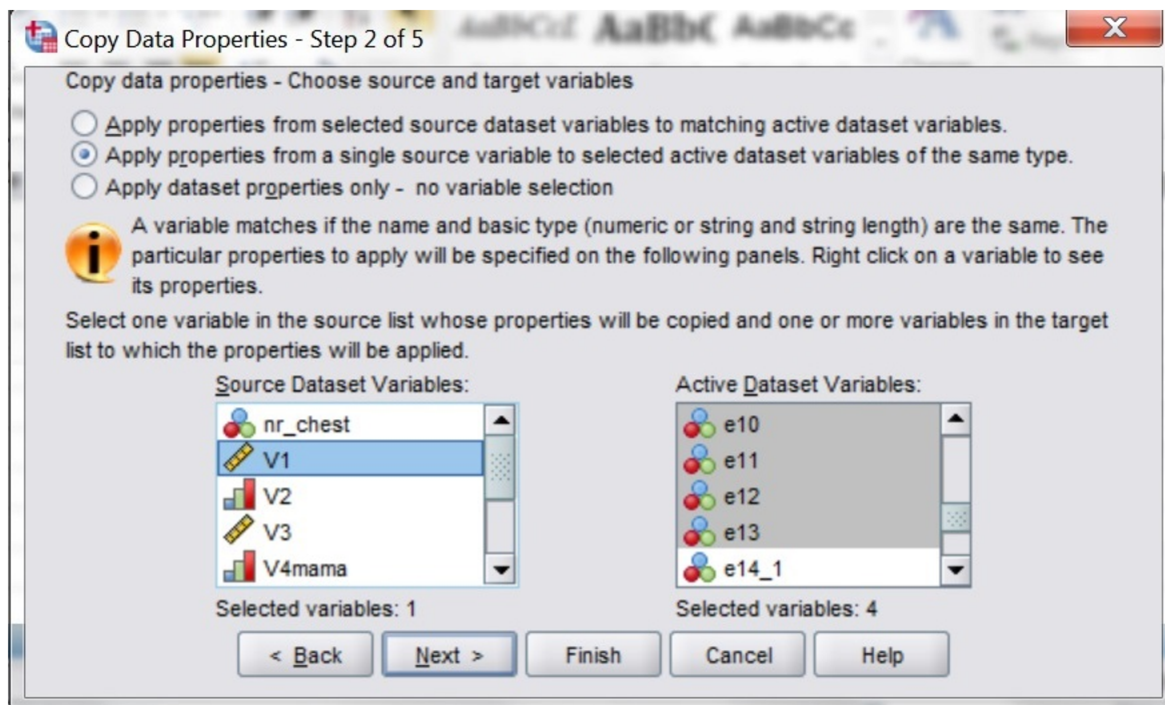
În graficul 6.16 am ales „Aplicarea proprietăților unei baze de date selectată asupra variabilelor care se potrivesc din baza de date activă". Două variabile se potrivesc dacă au același nume și același tip (numeric sau string; pentru variabilele „string" potrivirea necesită și același număr de caractere). Programul a identificat, în baza de date activă, două potriviri, astfel încât caracteristicile variabilelor „Tot_mem6cat" și „munc_strain", din fișierul extern, vor fi aplicate variabilelor cu același nume din „Active Dataset".

Dacă vrem ca în baza de date activă să fie create variabile cu structura celor din fisierul extern chiar dacă lipsesc variabilele pereche, alegem „Create matching variables in the active dataset if they do not already exist". În graficul 6.16, pe lângă cele două variabile perechi („Tot_mem6cat" și „munc_strain") au mai fost selectate și variabilele „V4tata" și V5 astfel încât numele și proprietățile acestora vor fi transferate în baza de date activă (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Următoarea opțiune constă în „Aplicarea proprietăților de la o singură variabilă sursă pentru variabilele de același tip selectate din fișierul activ".

În graficul 6.17 se poate vedea că în fișierul sursă (din care se copiază structurile datelor) a fost selectată variabila „V1" pentru ca proprietățile acesteia să fie aplicate asupra variabilelor „e10"-„e-13" (definite similar).

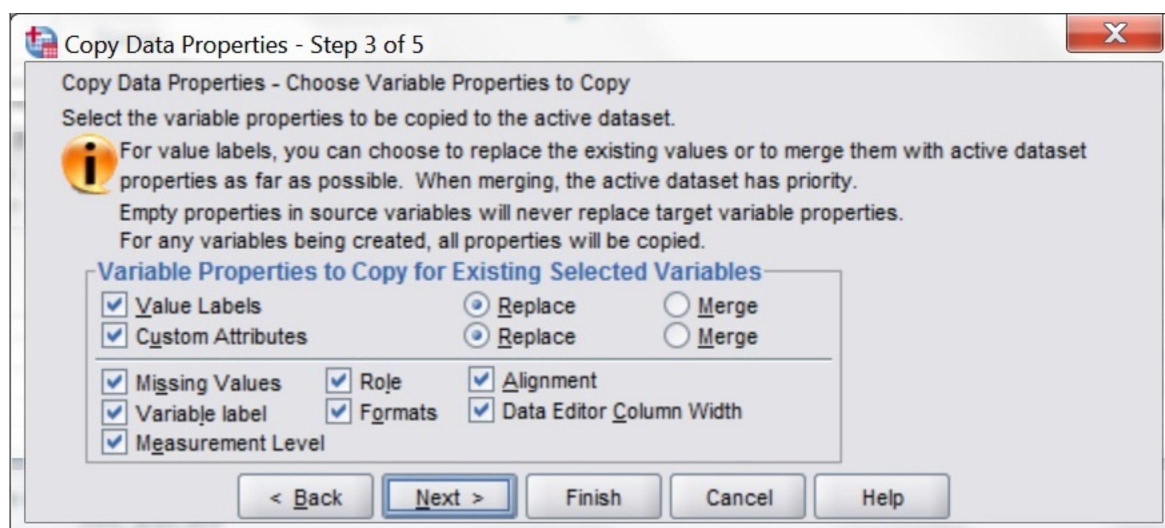
Graficul 6.17 Comanda Copy Data Properties (3) în SPSS



Tot în graficul 6.17 se poate opta pentru „Aplicarea doar a proprietăților fișierului tip bază de date, fără a selecta variabile" Proprietățile fișierelor includ documente, etichete ale fișierelor, seturi de răspunsuri multiple, seturi de variabile și modalități de ponderare a datelor.

Pentru fereastrele de dialog ilustrate în geaficele 6.16 și 6.17 (ambele vizând aplicarea proprietăților variabilelor din baze de date externe asupra bazei de date activă), următorul pas presupune a stabili proprietățile ce trebuie copiate. Graficul 6.18 ilustrează posibilele opțiuni (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 6.18 Comanda Copy Data Properties (4) în SPSS



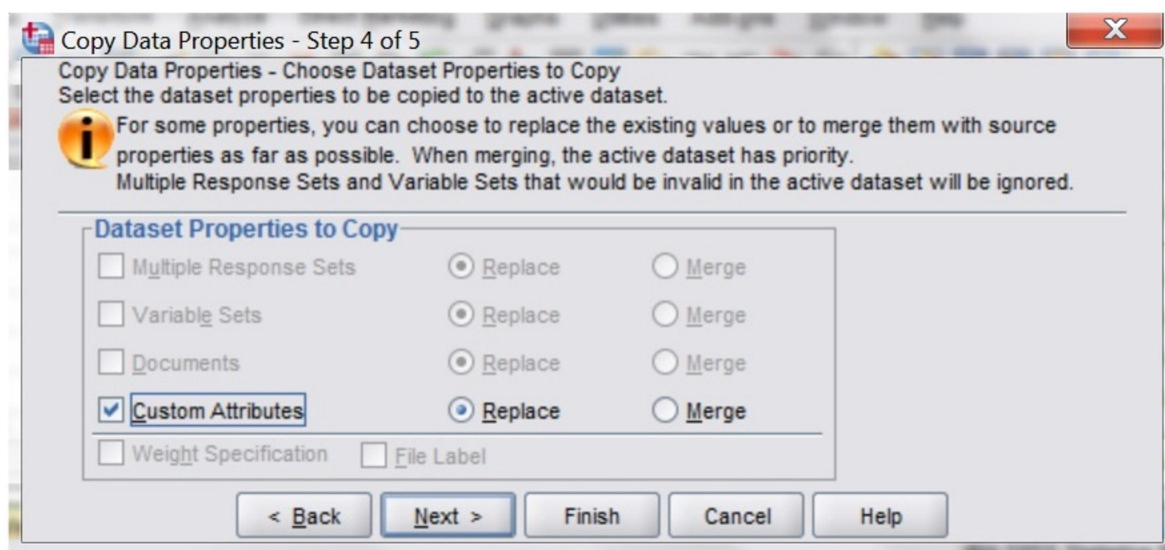
În partea de jos a ferestrei se poate opta pentru copierea datelor lipsă, a etichetelor variabilelor, a nivelulului de măsurare al variabilei, a rolulului acesteia (input, output, dublu rol, partiționare și divizare a bazei de date), a formatului variabilei, a alinierii valorilor în coloanele cu date și a lățimii coloanelor.

Pentru etichetele valorilor variabilelor și pentru atributele variabilelor întroduse în plus față de cele standard (Custom atributes) se poate opta pentru copierea din baza de date externă în cea activă sau se poate alege unirea (merge) caracteristicilor celor două baze de date ori de câte ori este posibil. În acest proces de unire (denumit „merge") proprietățile variabilelor din baza de date activă vor avea prioritate.

De notat, după cum se precizează în graficul 6.18, că proprietățile care nu corespund unor variabile în baza de date externă nu vor înlocui niciodată propritățile variabilelor țintă din baza de date activă. Pe de altă parte, pentru orice variabilă nou-creată în baza de date activă, toate proprietățile din baza de date externă vor fi copiate.

Următorul pas al demersului de copiere a caracteristicilor structurale ale variabilelor dintr-o bază de date externă în cea activă îl constituie, de fapt, copierea caracteristicilor globale ale bazei de date externă în cea internă. Se poate copia, de exemplu, modul de definire a seturilor de răspunsuri la întrebările cu alegere multiplă, definirea grupării variabilelor (Variable Sets), atributele suplimentare definite pentru variabilele din baza de date (care sunt vizibile în „Data Editor", „Variable View").

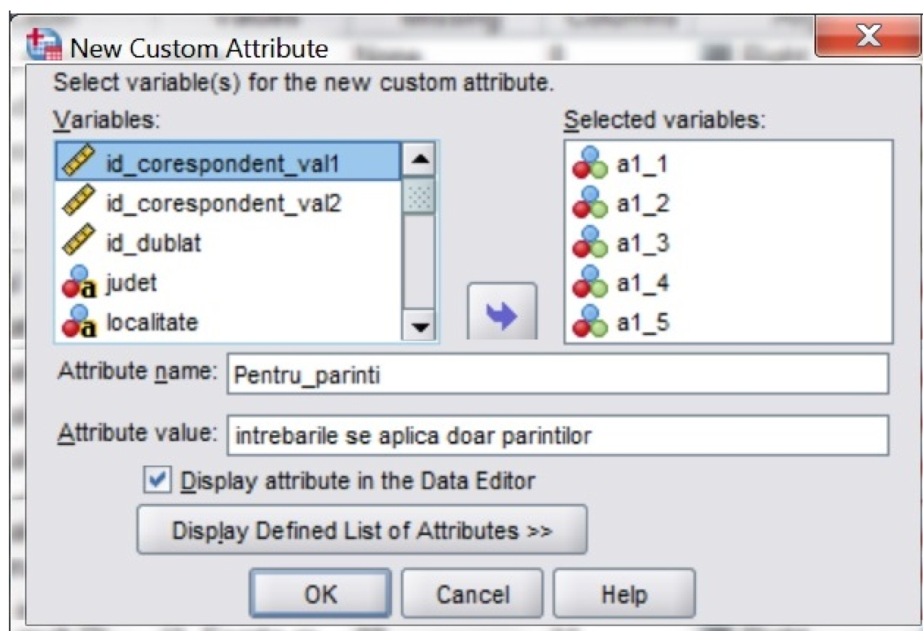
Graficul 6.19 Comanda Copy Data Properties (5) în SPSS



În plus, se mai poate copia structura variabilei care a fost definită pentru ponderarea bazei de date externă sau eticheta atribuită respectivei baze de date. Dând iarăși click pe butonl „Next" se va deschide ferastra finală în care vom fi informați despre numărul de variabile asupra cărora au fost aplicate proprietățile bazei de date externă și privitor la numărul de variabile copiate din respectiva bază de date. E precizat, e asemenea, câte proprietăți globale ale bazei de date au fost copiate. În final, se poate opta pentru lansarea comenzii de copierie a proprietăților sau pentru salvarea comenzilor într-o fereastră de tip sintaxă (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Următoarea opțiune a meniului „Data" se referă la definirea unor noi proprietăți ale variabilelorfolosindcomenzile **Data**, **New Custom Attribute**.

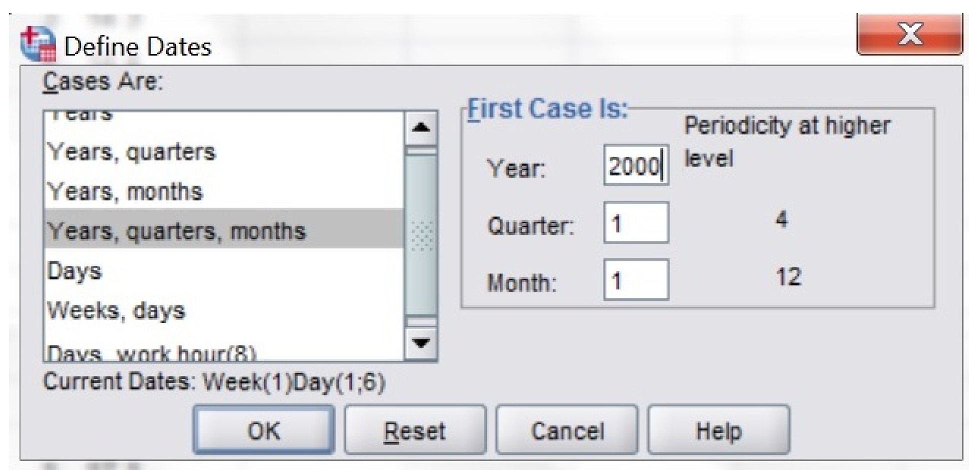
Graficul 6.20 Data, New Custom Attribute, în SPSS



În partea din stânga-sus a graficului 6.20 se pot selecta variabilele asupra cărora vor fi aplicate noile atribute. În exemplul nostru am selectat variabilele a1_1 - a1_5 și le-am deplasat în căsuța „Selected variables. Numele atributului nou introdus este „Pentru_parinti” iar valoarea acestuia este „intrebarile se aplică doar parintilor”. Fereastra permite setarea afișării noilor atribute în „Data Editor” precum și listarea tuturor atributelor definite (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Comanda **Data**, **Define Dates** permite definirea, în baza de date SPSS, a unor variabile care pot fi folosite pentru a stabili periodicitatea unei serii de timp și pentru a eticheta datele de ieșire din analiza seriilor de timp (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Graficul 6.21 Data, Define Dates, în SPSS



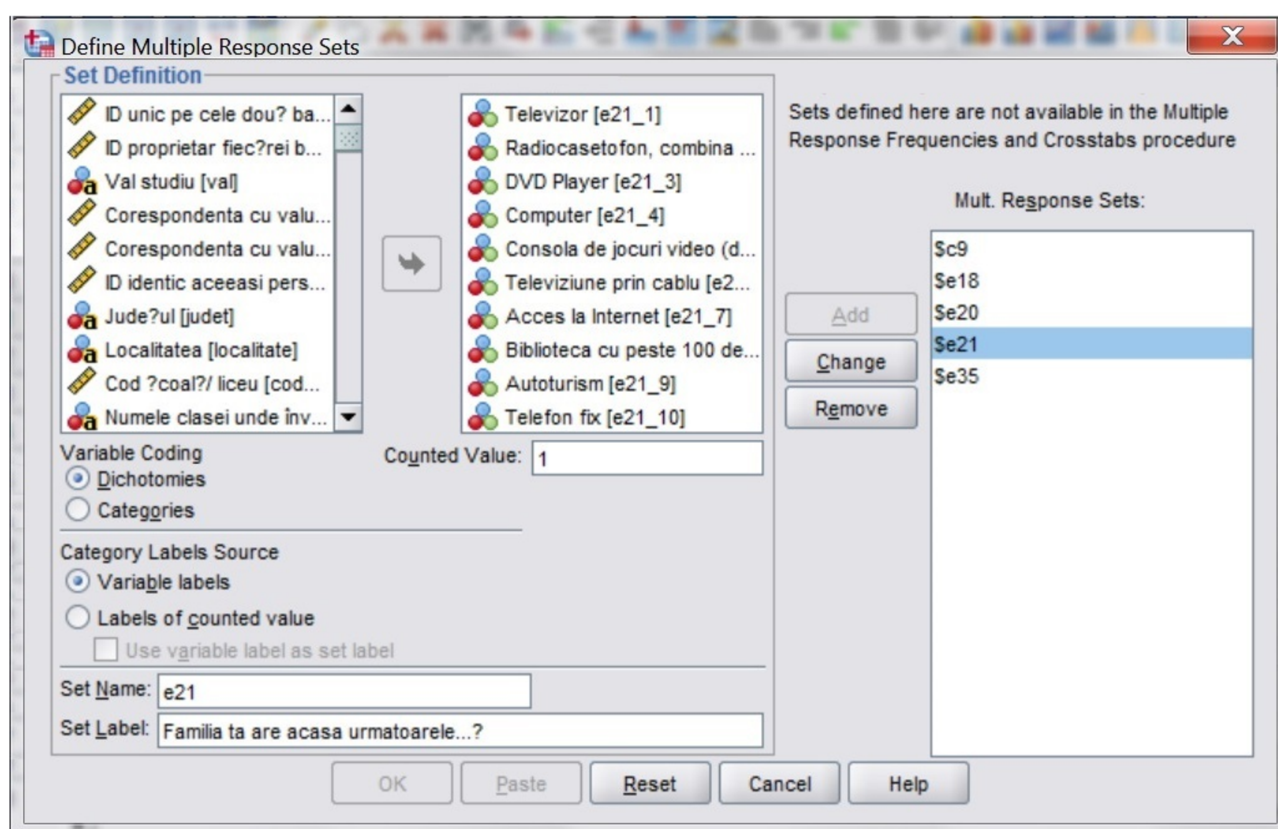
Fereastra 6.21 pezintă varianta de dată care e aplicată actualmente (numărul săptămânii pentru 6 zile lucrătoare) și cea nouă: anul (începând cu 2000), trimestrul și luna. Dând „OK” noua setare o va înlocui pe cea veche (IBM Knowledge Center, Home, SPSS Statistics 24.0).

6.5 Variantele DATA, Define Multiple Response, Validation

Fereastra de dialog deschisă prin comenzile „Data" „Define Multiple Response Set" permite gruparea variabilelor subsumate unor întrebări cu alegere multiplă. O situație tipică e cea a variabilelor dihotomice, care indică alegerile și respingerile subiectului pentru fiecare dintre variantele posibile.

În graficul 6.22 subiecții au fost rugați să indice care dotări se găsesc în gospodăria lor. Setul de răspunsuri a fost notat cu $e21, s-a specificat că e vorba de variabile dihotomice (Variable Coding / Dichotomies) și s-a specificat valoarea 1 („Counted value", indicând prezența aspectului vizat) ca fiind numărată pentru cele 10 variabile proprii fiecărei entități statistice din baza de date (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Graficul 6.22 Data Define Multiple Response Set in SPSS

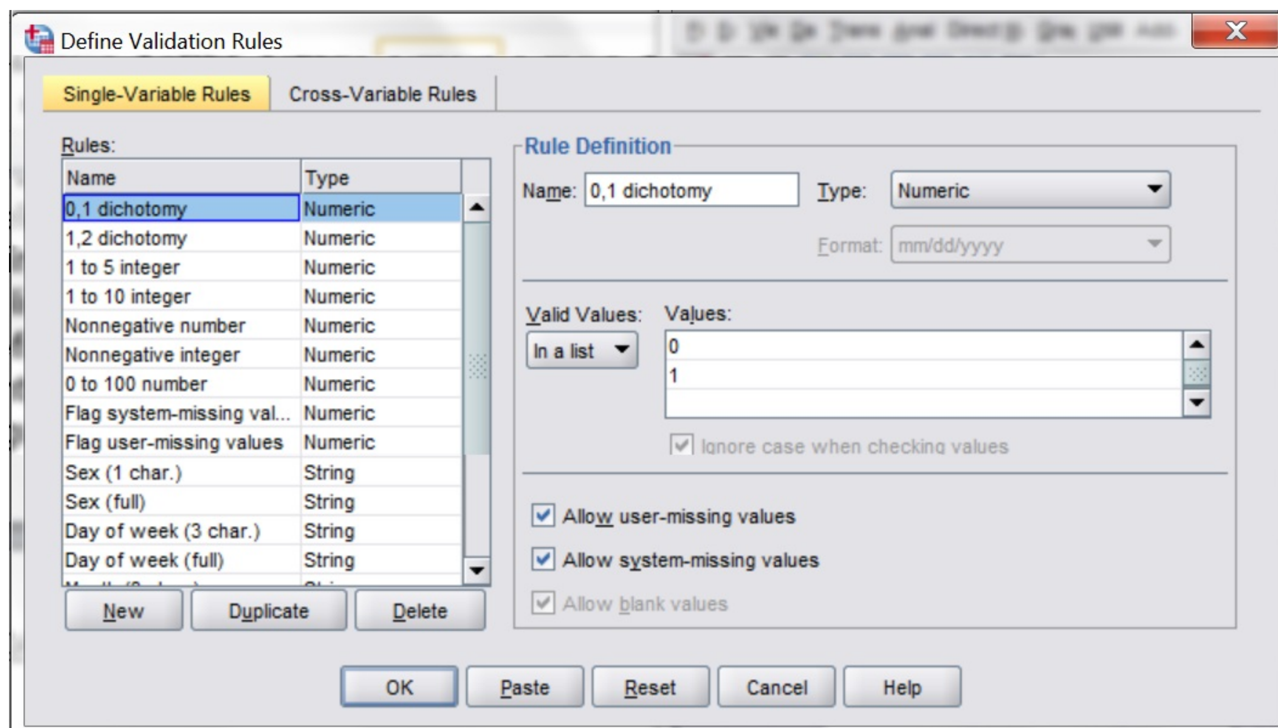


Pot fi definite și seturi de răspunsuri multiple pentru variabile cu mai multe categorii, dacă s-a formulat o întrebare cu mai multe variante de răspuns, dintre care se solicita doar un număr precis de alegeri. Evident, numărul de variante poate fi foarte mare în cazul unor variabile vizând, de exemplu, primele 3 orașe în care i-ar place unui tânăr să urmeze studiile superioare. În aceste condiții vor fi alese primele, să zicem, 20 cele mai populare orașe (adăugând și varianta „alt oraș") iar în baza de date vor fi construite 3 variabile, câte una pentru fiecare alegere solicitată de cei care au construit chestonarul (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Fereastra „**Data Define Validation Rules**" permite crearea și vizualizarea regulilor de validare pentru o variabilă singulară („Single-Variable Rules") respectiv pentru variabile asociate („Cross-Variable Rules"). Crearea și vizualizarea regulilor de validare necesită pregătirea datelor.

În graficul 6.23 este prezentată fereastra de dialog care permite să fie definite, șterse, modificate sau create noi reguli de validare la nivel de variabilă individuală (Single-Variable Rules) sau în cazul variabilelor asociate (Cross-Variable Rules).

Graficul 6.23 Data, Validation, Define Rules in SPSS



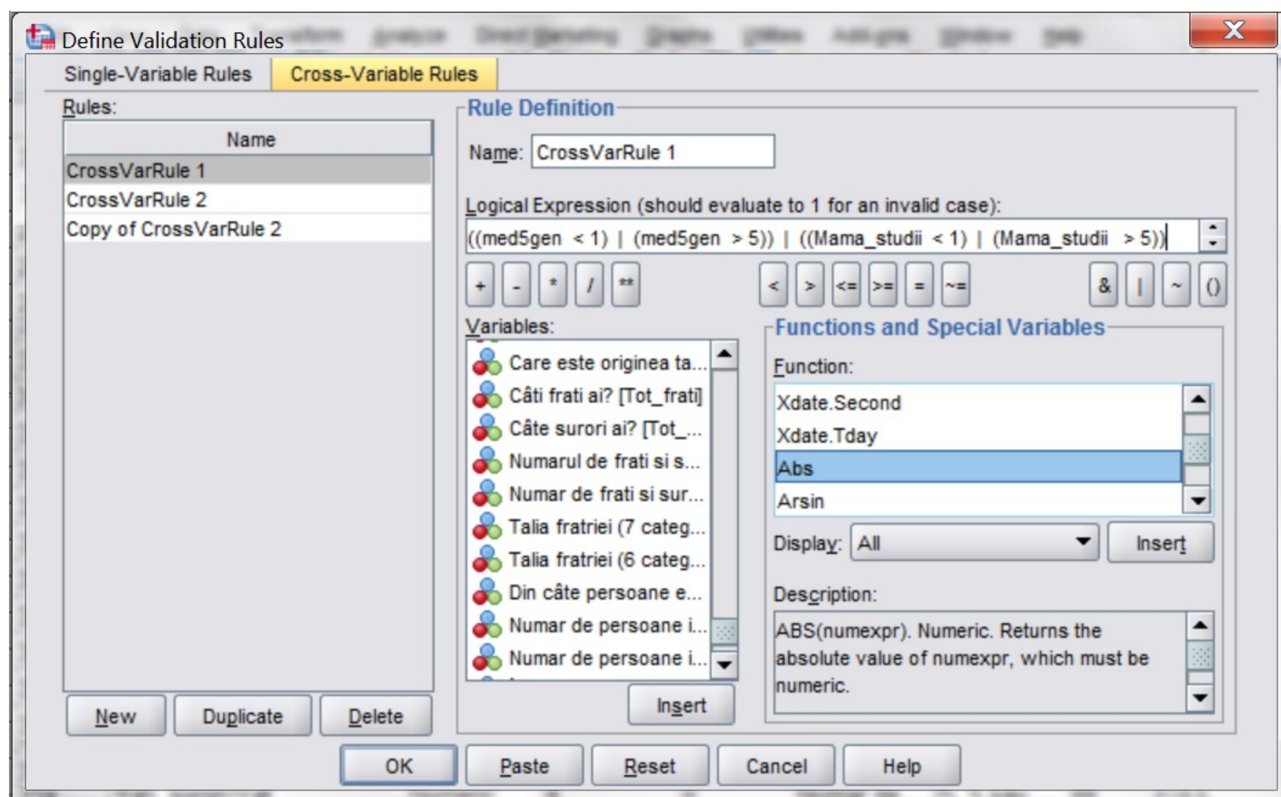
Butonul „New" din partea de jos a ferestrei permite definirea unei noi reguli de validare căreia i se poate atribui un nume („Name"), un anumit tip („Numeric", „String" sau „Date") și pentru care pot fi definite valorile valide într-o listă („in a list") sau într-un interval („Within a range"). Indiferent de varianta adoptată, se vor specifica valoarea minimă și cea maximă acceptate și, dacă e cazul, acceptarea valorilor fracționare. În partea de jos a ferestrei se poate stabili dacă sunt acceptate datele definite ca fiind lipsă de către utilizator („Allow user-missing values") precum și cele care lipsesc complet din baza de date („Allow system-missing values"). Butonul „Duplicate" va genera un set de reguli de validare identic cu unul deja existent în lista „Rules" (și selectat anterior cu click simplu) având adăugat la nume sintagma „Copy of". Cu ajutorul butonului „Delite" din partea de jos a graficului 6.23 putem să eliminăm regulile de validare de care nu mai avem nevoie. În final, prin butonul OK lansăm execuția comenzilor date anterior (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Fereastra „Data" „Validation" „Define Rules" „Cross-Variable rules" (a se vedea graficul 6.24) permite crearea, vizualizarea și modificarea regulilor de validare a asocierii dintre variabile. Expresia logică folosită tebuie să ia valoarea 1 pentru cazurile care nu sunt valide (și care trebuie excluse din analiză).

În exemplul din graficul 6.24, expresia logică devine 1 dacă membrul stâng al expresiei logice ia valoarea 1 sau dacă membrul drept ia valoarea 1. Membrul stâng va lua valoarea 1 dacă variabila „med5gen" va fi mai mică decât 1 sau mai mare decât 5 iar membrul drept devine 1 dacă variabila „mama_studii" are valori sub 1 sau peste 5.

Deci, e suficient ca una dintre condițiile logice să fie îndeplimită pentru ca, prin regula de validare (în exemplul nosru „CrossVarRule1”), să fie elimnate cazurile aflate în situația respectivă.

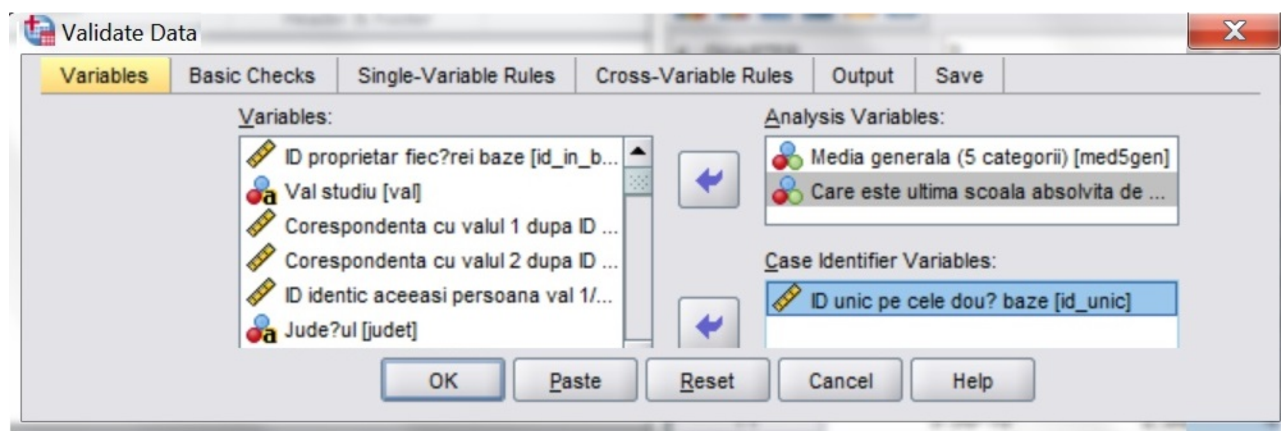
Grafic 6.24 Data, Validation, Define Rules, Cross-Variable Rules în SPSS



De notat faptul că în expresia logică („Logical expression”) pot fi folosite și funcțiile care sunt listate în partea dreaptă a graficului 6.24 (și sunt descrise în partea de jos a acestuia. În final se lansează comanda prin butonul „OK” (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Comanda **Data Validation Validate Data** se referă la utilizarea criteriilor de validare a datelor dar și la definirea unor noi criterii.

Graficul 6.25 Data Validation, Validate Data, Variables în SPSS



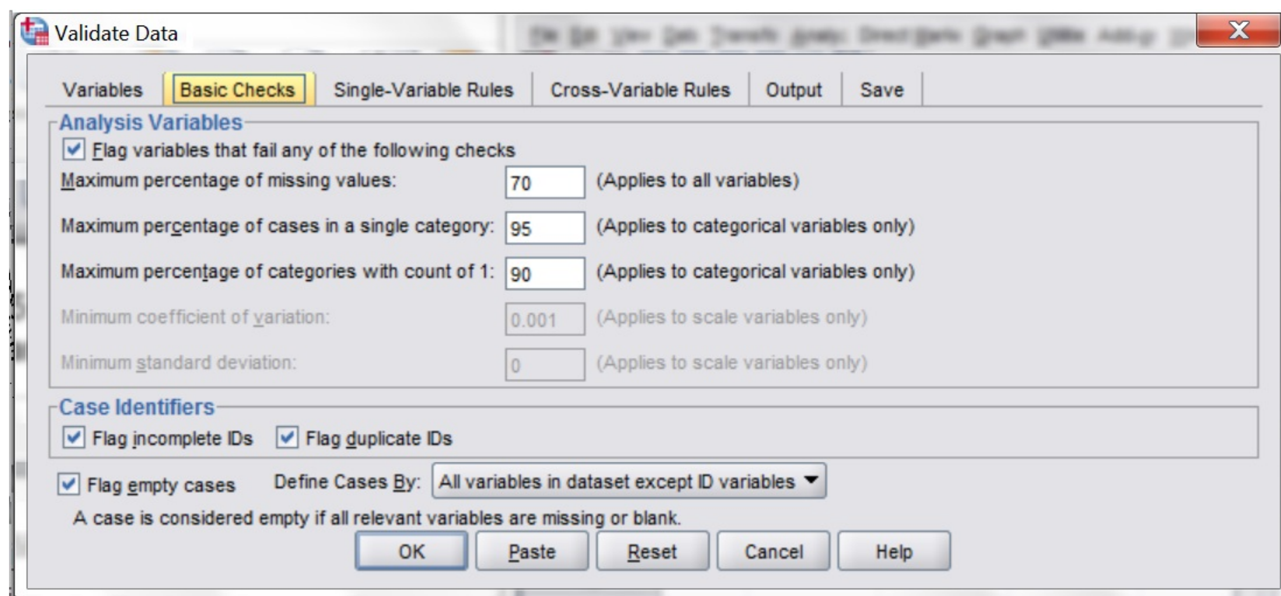
În fereastra 6.25 putem alege una sau mai multe variabile care să fie analizate precum și variabila utilizată drept identificator al cazurilor.

Variabila indicator al cazurilor numerotează cazurile din baza de date și permite identificarea dublurilor sau a valorilor de identificare incomplete.

Următoarea opțiune a aceleiași ferestre (Data Validation, Validate Data) este „Basic Chechs” care va indica toate variabilele care nu au îndeplinit una sau mai multe din următoarele cerințe:

- Procentajul maxim de date lipsă ____ (Aplicabil la toate variabilele)
- Procentaj maxim într-o categorie ____ (Aplicabil doar datelor categoriale)
- Maxim % categorii cu frevența 1 ____ (Aplicabil doar datelor categoriale)
- Coeficientul de variație minim ____ (Aplicabil doar datelor cantitative)
- Abaterea standard minimă ____ (Aplicabil doar datelor cantitative)

Graficul 6.26 Data Validation, Validate Data, Basic Checks, în SPSS



În graficul 6.26, în căsuța „Case Identifiers” se poate stabili care cazuri să fie marcate în baza de date: cazurile cu date de identificare incomplete și / sau cazurile pentru care s-au găsit duplicate (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

În partea de jos a graficului 6.26, se poate opta pentru marcarea cazurilor pentru care lipsesc datele (Flag empty cases) adică cele care pentru toate variabilele relevante înregistrează date lipsă sau „Missing system”. Din aceeași parta a graficului 6.26 se poate stabili dacă la definirea unui caz să participe toate variabilele din baza de date sau toate variabilele exceptându-le pe cea sau cele care furnizează identificarea cazurilor.

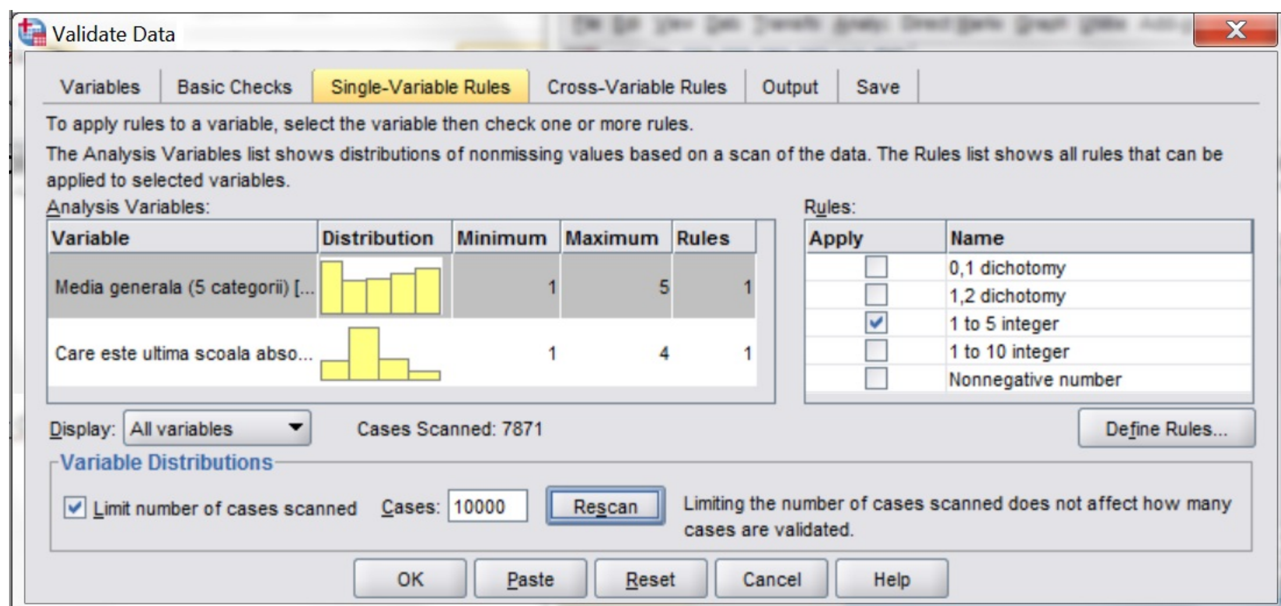
Următoarea opțiune pentru comanda „Data Validation” „Validate Data” este „Single variable Rules” și se referă, evident, la criteriile de validare a cazurilor pentru variabilele considerate în mod independent.

În graficul 6.27 e vizualizat modul de definire a regulilor de validare a cazurilor pentru variabilele considerate separate unele de celelalte.

În partea stângă a graficului apar rezultatele scanării variabilelor care au fost alese inițial („Media generală - 5 categorii” și „Care este ultima școală absolvită de mama dvs?”), inclusiv numărul cazurilor scanate. Numărul maxim de cazuri scanate se poate limita și se poate cere re-scanarea datelor (IBM Knowledge Center, Home, SPSS Statistics 24.0) .

Dacă dorim să aplicăm una dintre regulile pre-definite asupra unei variabie din lista existentă în stânga graficului 6.27 este suficient să o selectăm și apoi să dăm click pe regula aleasă („Rules:", „Apply").

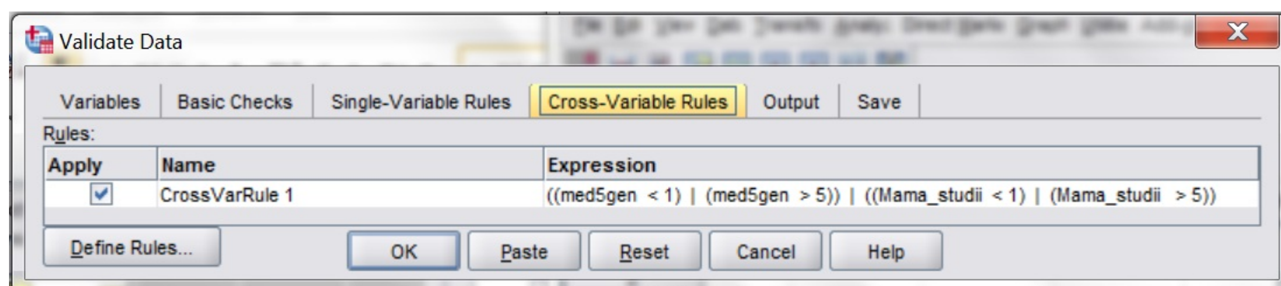
Graficul 6.27 Data Validation, Validate Data, Single variable Rules, în SPSS



Atragem atenția că pentru aceleași variaibile pot fi aplicate, dacă e nevoie, mai multe reguli de validare a cazurilor. În exemplul nostru concret, am aplicat regula ca variabila să aibă valori întregi între 1 și 5 atât pentru „Media generală" cât și pentru „Ultima scoală absolvită de mama dvs." În plus, butonul „Define Rules" permite definirea unor noi reguli de validare în mod identic cu fereastra prezentată în graficul 6.23 (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Opțiunea următoare din „Data Validation", „Validate Data" e cea care se referă la validarea cazurilor prin reguli aplicate simultan mai multor variabile.

Graficul 6.28 Data Validation, Validate Data, Cross Variable Rules, în SPSS



În graficul 6.28 s-a aplicat regula cu numele „CrossVarRule 1" având expresia prezentată în grafic (care devine adevărată dacă valorile variabilei „med5gen" respectiv „Mama_studii" nu se încadrează între valorile 1 și 5. Și această fereastră ne permite să definim noi reguli („Define rules", butonul de jos). Acestea vor fi definite într-o fereastră identică cu cea din graficul 6.24, obținută prin comenzile: „Data" „Validation" „Define Rules" „Cross-Variable Rules" (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Următoarea opțiune pentru „Validate Data”, care este „Output”, permite setarea datelor de ieșire vizând rezultatele validării.

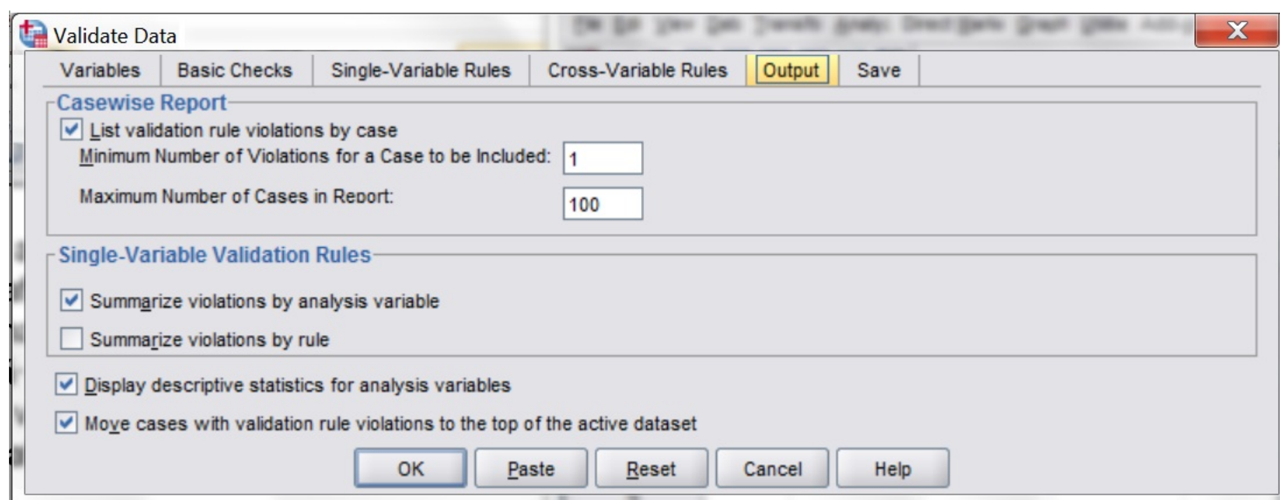
Dacă s-au aplicat reguli de validare cu o singură variabilă sau cu 2 variabile concomitent, se poate cere un raport conținând încălcări ale regulilor de validare pentru cazuri individuale (vezi graficul 6.28, „list validation rule by case”).

În cadrul aceleiași opțiuni putem stabili numărul minim de violări ale regulilor pe care trebuie să-l întrunească un caz pentru a fi cuprins în rapot și, de asemenea, numărul maxim de cazuri care pot fi cuprinse în respectivul raport; limita maximă admisă este de 1000 (IBM Knowledge Center, Home, SPSS Statistics 24.0).

„Single-Variable Validation Rules” da posibilitatea, în cazul în care au fost definite reguli de validare cu o singură variabilă, să se aleagă modul de afișare a rezultatelor. Există și posibilitatea de a renunța complet la afișarea rezultatelor mai sus menționate (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Dacă alegem varianta „Summarize violations by analysis variable” pentru fiecare variabilă de analiză, datele de ieșire vor prezenta toate regulile de validare cu o singură variabilă care au fost încălcate și frecvențele valorilor care încalcă fiecare regulă. De asemenea, este prezentat numărul total de încălcări ale regulilor cu o singură variabilă pentru fiecare variabilă.

Graficul 6.29 Data Validation, Validate Data, Output, în SPSS



Optând pentru „Summarize violations by rule”, pentru fiecare regulă de validare cu o singură variabilă vor fi afișate variabilele care au încălcat regula și frecvența valorilor nevalide per variabilă. În plus se, raportează numărul total de valori care încalcă fiecare regulă în cadrul variabilelor.

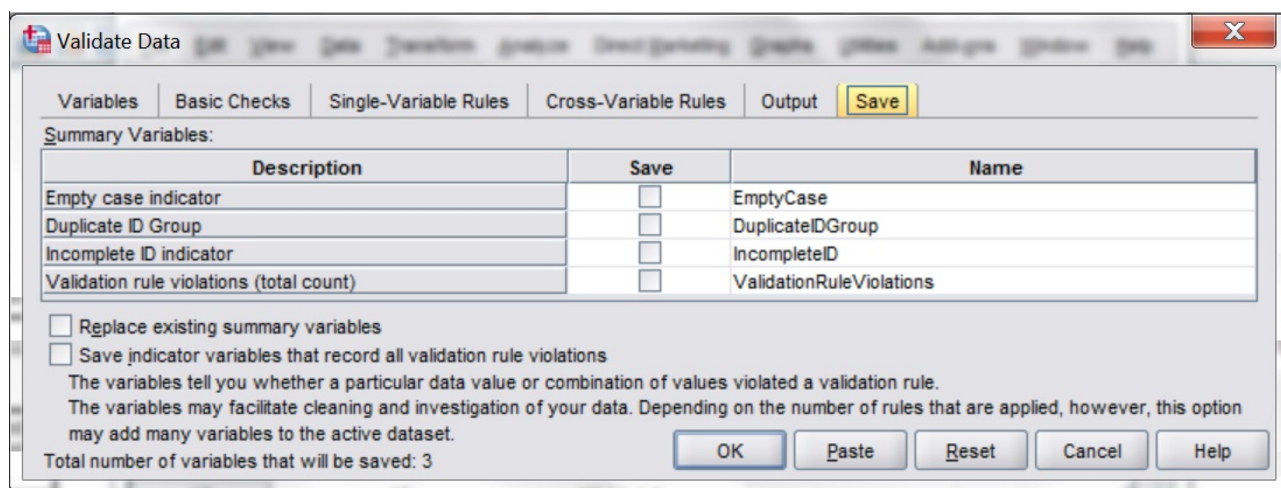
„Display descriptive statistics” permite afișarea statisticilor descriptive pentru variabilele de analiză. Pentru fiecare variabilă categorială este generat un tabel de frecvență. Un tabel statistic care include media, abaterea standard, minimul și maximul este generat pentru variabilele cantitative.

„Move cases with validation rule violations” permite mutarea cazurilor în care există încălcări ale regulilor în raport cu una sau mai multe variabilă în partea de sus a bazei de date activă, pentru o examinare mai ușoară (IBM Knowledge Center, Home, SPSS Statistics 24.0).

„Save”, ultima fereasrtă de dialog pentru „Validate Data”, permite să fie salvate, în baza de date activă, variabilele care surprind încălcarea regulilor.

„Summary Variables” se referă la variabilele individuale care pot fi salvate prin simpla bifare a unui buton din dreptul comenzii „Save”. Variabilele au nume implicite dar pot fi modificate.

Graficul 6.30 Data Validation, Validate Data, Save, în SPSS



Varianta „Empty case indicator” generează o variabilă ale cărei valori sunt 1 pentru cazurile vide și 0 în celelalte situații.

Alegând „Duplicate ID Group” Cazurile care au același identificator de caz (altul decât cazurile cu identificatori incompleți) vor avea același număr de grup. Cazurile cu identificatori unici sau incompleți sunt codificați 0.

Selectând „Incomplete ID indicator” se va genera o variabilă având valoarea 1 pentru identificatorii de caz pentru care informațiile lipsesc sau sunt incomplete și 0 în celelalte situații (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Dacă optăm pentru „Validation rule violations” în baza de date SPSS va fi construită o variabilă care va indica numărul total al încălcării regulilor formulate în raport cu o variabilă respectiv cu mai multe variabile simultan.

„Replace existing summary variables” permite ca variabilele salvate în baza de date să le înlocuiască pe cele preexistente (dacă numele coincid).

Prin „Save indicator variables that record all validation rule violations” generăm variabile care indică dacă unele valori ori combinații de valori încalcă regulile de validare. Fiecare variabilă corespunde aplicării unei reguli de validare și are valoarea 1 dacă respectivul caz încalcă regula și valoarea 0 dacă nu o încalcă. Variabilele construite în acest mod pot facilita „curățirea” bazei de date dar, dacă numărul lor e prea mare, pot îngreuna acest proces (Ibid).

6.6 Variantele DATA, Identify Duplicate Cases, Identify Unusual Cases, Sort cases în SPSS

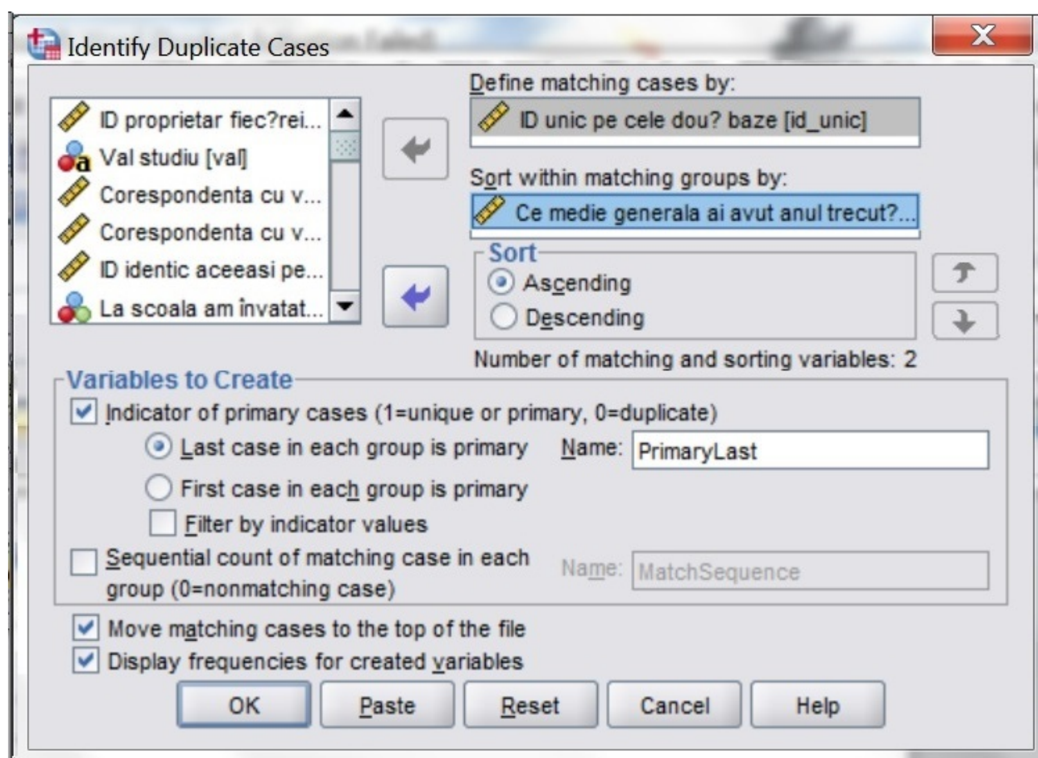
Următoarea opțiune a comenzii „Data” este „Identify Duplicate Cases” vizând identificarea cazurilor dublate în baza de date din diferite motive:

- erori datorate intoducerii de mai multe ori a aceluiași caz în baza de date
- cazuri care au o serie de identificări primare comune, de exemplu pentru că e vorba despre membrii unei familii care locuiesc în aceiași casă, însă au valori secundare de identificare diferite

- Cazuri multiple care reprezintă, de fapt, același caz dar cu identificări secundare diferite (aceeași persoană a făcut achiziții pentru firme diferite).

Opțiunea „Identify Duplicate Cases” permite definirea cazurilor dublate și oferă un anumit control în selectarea cazurilor valide față de cele dublate.

Graficul 6.31 Data, Identify Duplicate Cases, în SPSS



Căsuța „Define matching cases by” permite alegerea variabilelor după care se stabilesc dublurile. Cazurile sunt considerate duplicate dacă valorile coincid pentru toate variabilele selectate. Identificarea cazurilor care se potrivesc 100% în toate privințele presupune selecția tuturor variabilelor (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Cazurile sunt sortate automat în baza de date în funcție de variabilele utilizate pentru evaluarea potrivirii cazurilor. Căsuța „Sort within matching groups by” permite selectarea unor variabile suplimentare de sortare care vor determina ordinea cazurilor din fiecare grup de cazuri care sunt similare (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

- Pentru fiecare variabilă de sortare se poate alege ordonarea crescătoare sau descrescătoare.
- Dacă se aleg mai multe variabile de sortare, cazurile sunt sortate întâi după prima variabilă apoi după a doua, în interiorul categoriilor primeia. Apoi sortarea se face după a treia variabilă în interirul categoriilor definite de combinarea primelor două, etc.
- Butoanele cu săgeți ascendente și descendente din dreapta graficului 6.31 permit schimbarea ordinii de sortare a variabilelor
- Ordonarea în cadrul fiecărui grup de similaritate este importantă deoarece determină care caz se va afla în situața de „Primary last” sau „Primary First” (primul sau ultimul în grup) dintre care se vor alege cazurile valide.

Dacă dorim să filtrăm / eliminăm toate cazurile din grupurile de similaritate exceptând pe cei cu vechimea în muncă cea mai mare, aceea variabilă o vom folosi în căsuța de dialog „Sort within matching groups by”.

„Indicator of primary cases” generearză o variabilă cu valoarea 1 pentru toate cazurile unice și cele identificate ca fiind cazul primar în fiecare grup de similaritate și valoarea 0 pentru duplicatele non-primare din fiecare grup.

„Primary case” poate fi primul ori ultimul caz din fiecare grup, în funcție de ordinea de sortare din grupul de similaritate. Dacă lipsește variabila de sortare, se va păstra, în fiecare grup, ordinea inițială a fișierului (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Opțiunea „Filter by indicator values” permite utilizarea variabilei care indică „primary cases” ca filtru pentru a exclude din rapoarte și analize duplicatele care nu sunt „primary” fără a șterge acele cazuri din baza de date.

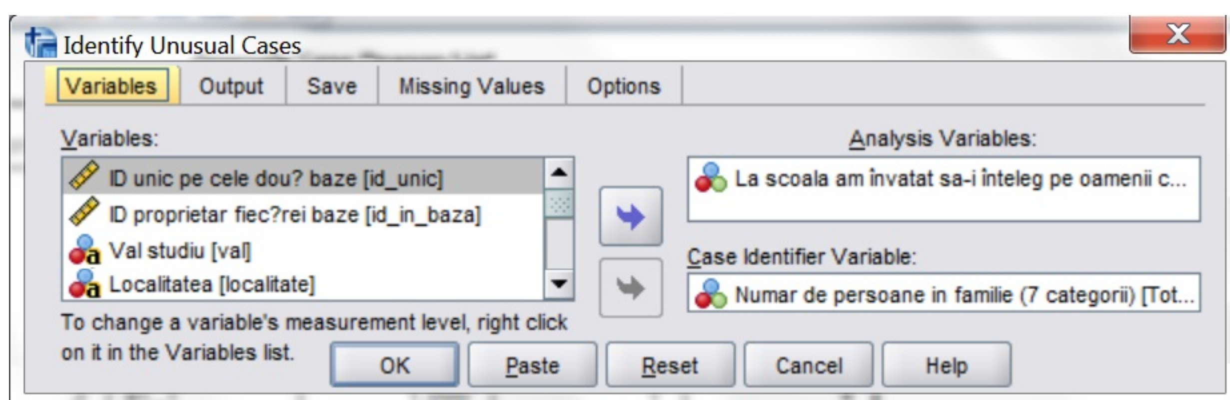
Opțiunea „Sequential count of matching cases in each group” creează o variabilă care ia valori de la 1 la N pentru cazurile din fiecare grup de similaritate. Atribuirea valorilor se bazează pe ordinea cazurilor din fiecare grup (fie ordinea inițială a fișierului, fie cea determinată de variabilele de sortare).

Opțiunea „Move matching cases to the top” determnă sortarea bazei de date astfel încât toate grupurile de similaritate să se afle în partea de sus a acesteia, ușurând vizualizarea cazurile de potrivire din „Data Editor” (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Dacă selectăm „Display frequencies for created variables” vor fi afișate, în fișierul „Output”, distribuțiile de frecvență pentru toate variabilele nou-create. De exemplu, pentru o variabilă „Primary Indicator”, tabelul „Output” va indica numărul de cazuri duplicat cuantificat cu 0, respectiv numărul de cazuri unice sau „Primary” notate cu 1.

Următoarea opțiune a meniului „Data” este „Identify Unusual Cases”, menită să identifice cazuri neobișnuite pe baza abaterilor de la normele de grup ale cazurilor. Analiza e folosită în faza exploratorie a cercetării.

Graficul 6.32 Data, Identify Unusual Cases, Variables, în SPSS



Algoritmul e conceput pentru a detecta anomalii generice, și nu pentru aplicații specifice, cum ar fi asistența medicală sau finanțele. Procedura de identificare a cazurilor neobișnuite găsește și raportează aceste valori, astfel încât analistul să poată decide gestionarea lor. De notat că, de obicei, cazurile neobișnuite provin din înregistrări unice, nu din erori de introducere a datelor (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Procedura definește grupuri de similaritate, norme de grup atât în cazul variabilelor continue cât și al celor categoriale, indicatori de anomalie bazați pe abateri de la normele grupului de similaritate și valori de impact depinzând de măsura în care variabilele contribuie la un caz considerat neobișnuit.

În baza de date fiecare rând constituie un caz distinct și fiecare coloană reprezintă o variabilă specifică pe care se bazează grupurile de similaritate. O variabilă de identificare a cazurilor poate fi disponibilă în baza de date pentru marcarea rezultatelor, dar aceasta nu va fi utilizată în analiză. Valorile lipsă sunt permise dar cele destinate ponderării cazurilor vor fi ignorate.

Este important faptul că soluția poate depinde de ordinea cazurilor iar minimizarea acestui efect se poate face prin ordonarea aleatoare a cazurilor Observație: se poate testa stabilitatea soluției utilizând mai multe sortări aleatoare (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Ipoteze: algoritmul presupune că toate variabilele sunt independent, că nici una nu e constantă și că nu există date lipsă pentru variabilele de intrare. Fiecare variabilă continuă e presupusă a avea o distribuție normală iar fiecare variabilă categorială este considerată a avea o distribuție multinomială.

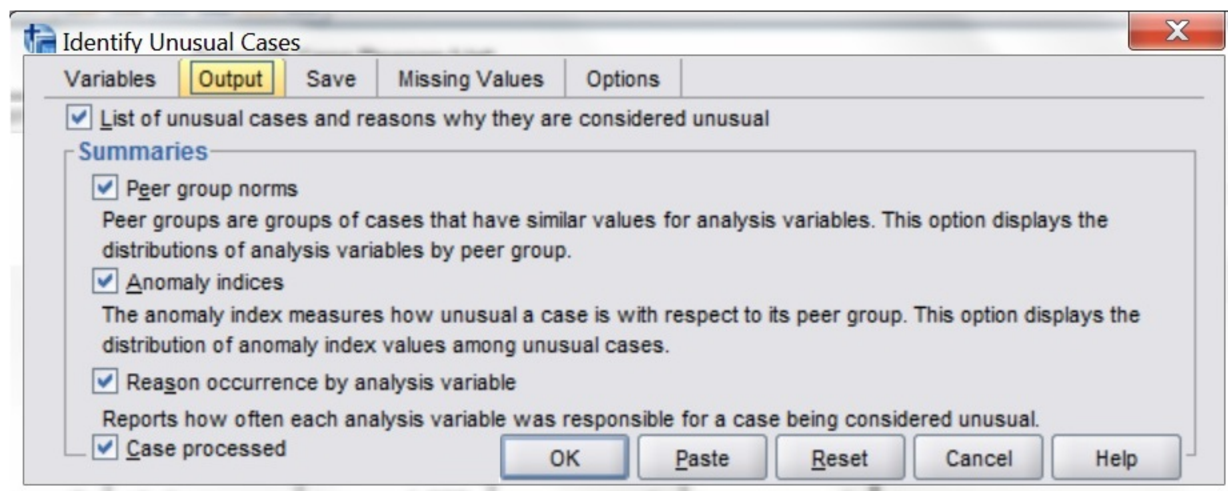
Se recomandă analiza de corelație pentru a testa independența a oricăror două variabile continue și tabelele de asociere pentru a testa independenței tuturor perechilor de variabile categoriale. Procedura „Means" testează independența între o variabilă continuă și una categorială. Utilizând analiza „Explore" se poate testa normalitatea unei variabile continue. Testul Hi pătrat permite testarea faptului că o variabilă categorială are distribuția multinomială specificată (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Pentru lansarea procedurii se vor alege una sau mai multe variabile de analiză și, opțional, o variabilă identificator a datelor de ieșire (etichetă). De notat că variabilele care intră în analiză trebuie să aibă definit în mod corect nivelul de măsurare deoare acesta influențează rezultatele analizei (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Selectând „List of unusual cases and reasons why they are considered unusual" vom genera trei tabele:

- Lista identificatorilor cazurilor neobișnuite (afișează cazuri identificate ca neobișnuite și prezintă valorile corespunzătoare ale indicelui de anomalie).

Graficul 6.33 Data, Identify Unusual Cases, Output, în SPSS



- Lista de identificare a grupurilor de similaritate din care fac parte cazurile neobișnuite și informații referitoare la respectivele grupuri
- Lista motivelor de anomalie, care afișează numărul cazului, variabila care reprezintă motivul anomaliei, valoarea de impact a variabilei, valoarea variabilei și norma variabilei pentru fiecare motiv.

De notat faptul că toate tabelele sunt ordonate după indicii de anomalie în ordine descrescătoare. În plus, valorile de identificare ale cazurilor sunt afișate dacă variabila identificator de caz e menționată (Case idetifier variable).

Alegând „Peer group norms” vom determina afișarea tabelului normelor variabilei continue (dacă există) și tabelul normelor categoriale. Tabelul cu normele variabilei continue afișează media și abaterea standard a fiecărei variabile continue pentru fiecare grup de similaritate. Tabelul cu normele variabilelor categoriale afișează valoarea modală, frecvența și procentul tuturor variabilelor în cazul fiecăriu grup de similaritate. Media variabilelor continue și valoarea modală a celor categoriale sunt folosite ca valori normale în analiză (IBM Knowledge Center, Home, SPSS Statistics 24.0).

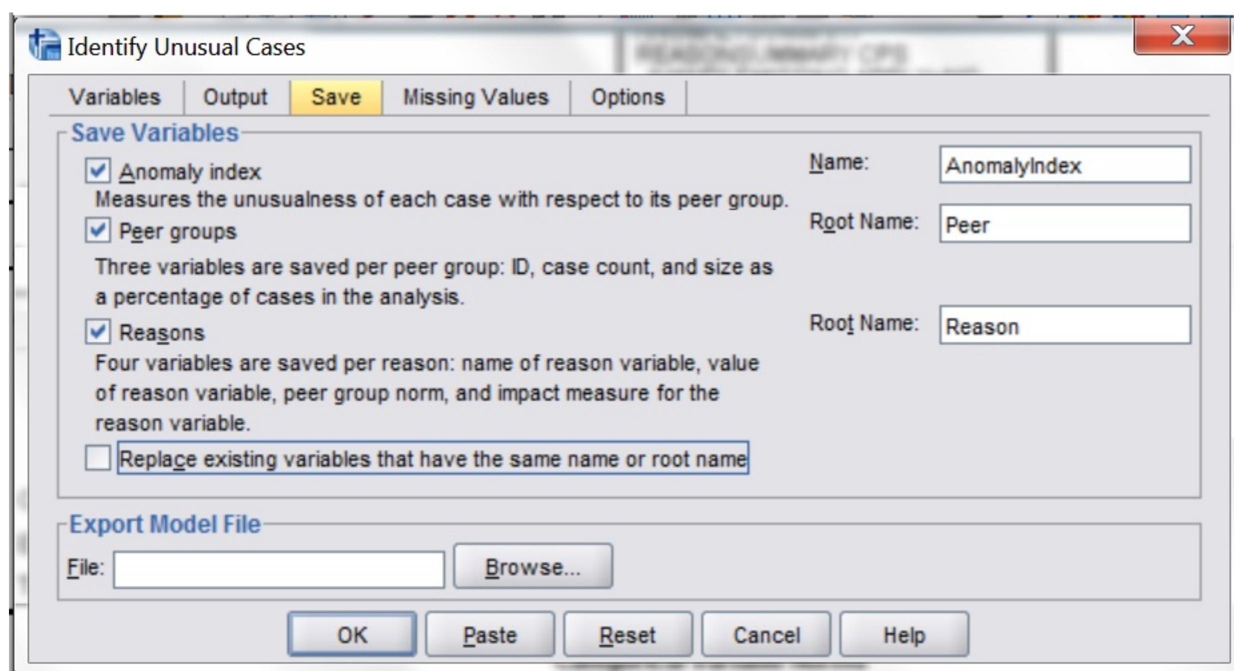
Selectând „Anomaly indices” se va afișa rezumatul indicelui de anomalie conținând statistici descriptive ale cazurilor considerate cele mai neobișnuite.

Optând pentru „Reason occurrence by analysis variable”, tabelul output va afișa frecvența și procentele apariției fiecărei variabile ca motiv. Tabelul redă și statisticile descriptive ale impactului fiecărei variabile. Dacă maximul de motive e setat la 0 în „Opțions”, acest tabel nu va fi disponibil.

„Cases processed” afișează un rezumat al analizei cazurilor conținând frecvența și ponderile pentru toate cazurile din baza de date, cazurile incluse și excluse din analiză și apartenența acestora la fiecare grup de similaritate (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Fereastra de dialog „Data, Identify Unusual Cases, Save” permite salvarea, în baza de date, a variabilelor definite anterior pentru a controla grupurile de cazuri (eventual înlocuind variabilele anterioare cu nume identic).

Graficul 6.34 Data, Identify Unusual Cases, Save, în SPSS



„Anomaly index”, care măsoară cât de neobișnuit e fiecare caz față de propriul grup de similaritate, salvează valoarea indicelui de anormalitate pentru fiecare caz într-o variabilă cu numele specificat.

Selectând „Peer groups”, trei variabile sunt salvate pentru grupul de similaritate: ID-ul grupurilor, numărul cazului și mărimea ca procent al cazurilor din analiză. De exemplu, dacă se specifică numele de rădăcină Peer, variabilele Peerid, PeerSize și PeerPctSize vo fi generate. „Peerid” e ID-ul peer-grupului unde se află cazul, „PeerSize” este dimensiunea grupului, iar „PeerPctSize” este dimensiunea grupului ca procent.

Alegând „Reasons”, patru variabile vor fi salvate pentru fiecare motiv: numele variabilei „motiv”, valorile variabilei „motiv”, normele grupurilor de similaritate și impactul măsurat pentru variabilele motiv.

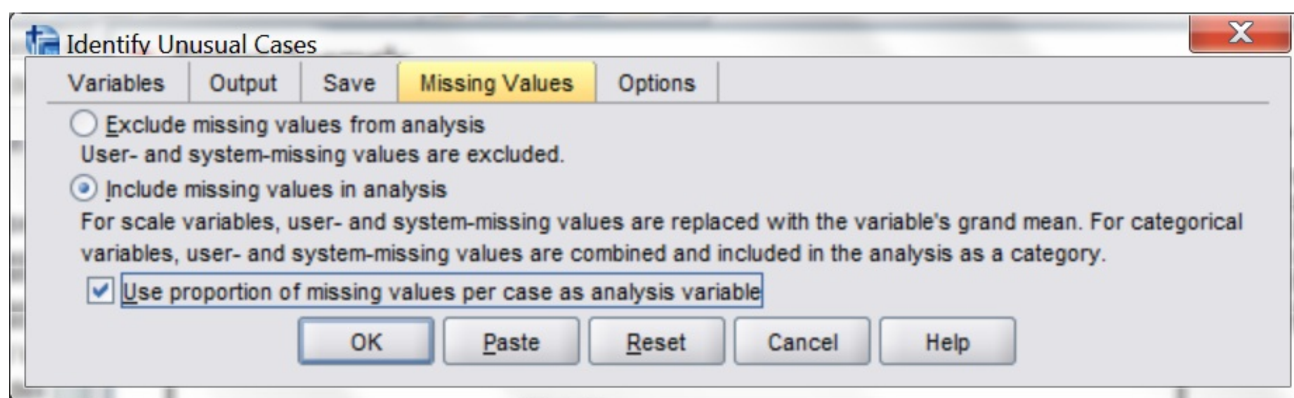
Evident, opțiunea nu e disponibilă dacă numărul motivelor a fost setat 0 în „Options” (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

În partea de jos a ferestrei din graficul 6.34 se poate opta pentru înlocuirea variabilelor existente care au același nume sau aceeași rădăcină.

Selectarea opțiunii „Export Model File” permite salvarea modelului de analiză a datelor neobișnuite în format XML (conținând comnzile SPSS).

Următoarea fereastră proprie opțiuneii „Data Identify unusual cases” este „Missing Values”

Graficul 6.35 Data Identify unusual cases, Missing Values in SPSS



Fereastra de dialog „Missing Values” e folosită pentru a controla modul de gestionare a datelor lipsă definite de utilizator și a celor „system-missing”.

Optând pentru „Exclude missing values from analysis” vom exclude complet din analiză datele lipsă, indiferent de natura acestora.

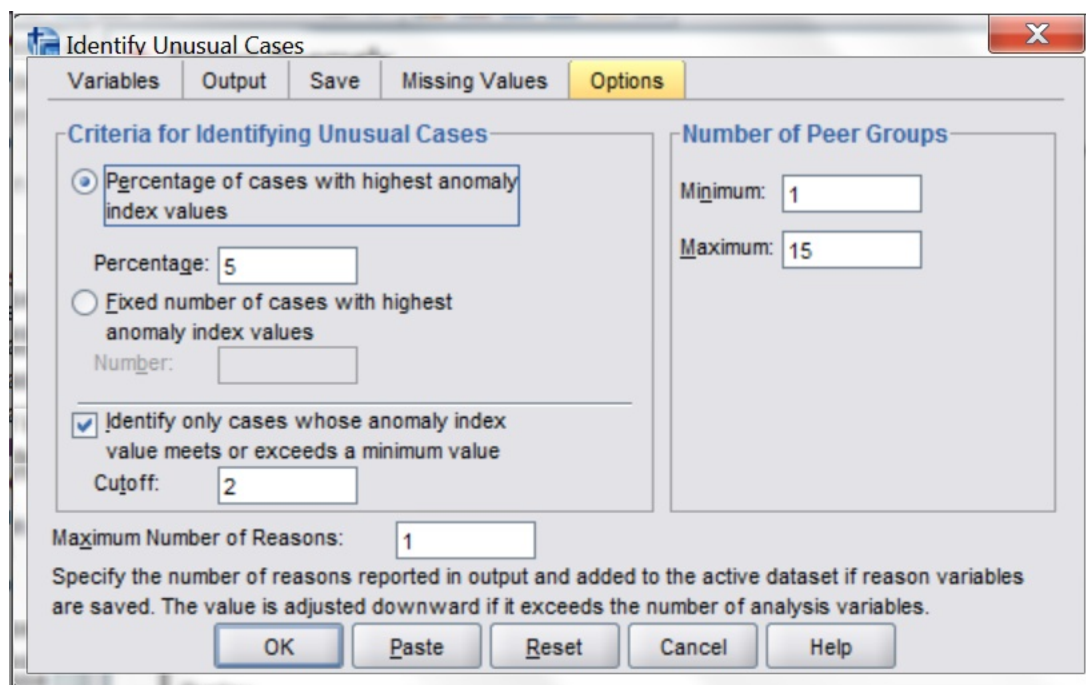
Dacă preferăm varianta „Include missing values in analysis”, valorile lipsă ale variabilelor cantitative continue sunt înlocuite cu media iar valorile declarate lipsă pentru variabilele categoriale sunt grupate și tratate ca o categorie validă (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Bifând „Use proportion of missing values per case as analysis variable” se solicită crearea unei variabile suplimentare care să reprezinte proporția variabilelor pentru care nu există date valide în fiecare caz din baza de date și utilizarea acesteia în analiză.

Ultima fereastră pentru „Data Identify unusual cases” este „Options” și dă posibilitatea stabilirii unor criterii pentru identificarea cazurilor neobișnuite și a setării numărului maxim respectiv minim al grupurilor de similaritate.

Setările cuprinse în „Criteria for Identifying Unusual Cases" permit să alegem numărul cazurilor considerate neobișnuite. Putem stabili procentajul cazurilor neobișnuite („Percentage of cases with highest anomaly index values") sau putem alege numărul exact de cazuri „anormale" („Fixed number of cases with highest anomaly index values"). De asemenea, avem posibilitatea să cerem selecția cazurilor neobișnuite al căror indice de anormalitate atinge sau depășește o anumită valoare alegând varianta „Identify only cases whose anomaly index value meets or exceeds a minimum value" (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 6.36 Data Identify unusual cases, Options in SPSS



În partea dreaptă a ferestrei putem indica numărul de grupuri de similaritate din analiză („Number of Peer Groups"). Procedura va căuta cel mai bun număr de grupuri de similaritate între valorile specificate (evident, valoarea minimă nu trebuie să depășească valoarea maximă iar dacă valorile specificate sunt egale, procedura presupune un număr fix de grupuri de similaritate). De notat că dacă numărul de grupuri de similatitate care pot fi suportate de datele concrete e inferior celui specificat, procedura poate produce mai puține grupuri .

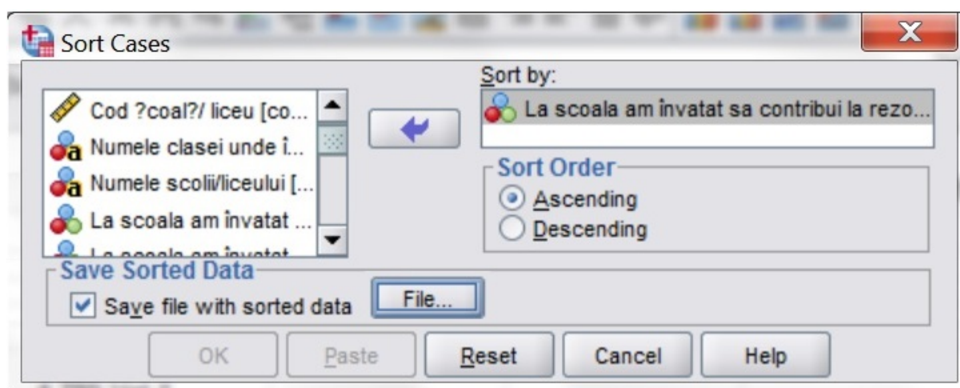
Opțiunea „Maximum Number of Reasons" stabilește numărul maxim de „motive" luat în calcul în analiza cazurilor neobișnuite. Un motiv constă în măsurarea impactului variabilei, indicarea numelui acestei variabile, a valorii variabilei și a valorii grupului de similaritate corespunzător. Dacă valoarea specificată pentru numărul maxim de motive este egală cu sau depășește numărul de variabile utilizate în analiză, vor fi afișate toate variabilele (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Următoarea opțiune a instrucțiunii **Data** este **Sort cases** și permite ordonarea bazei de date după una sau mai multe variabile.

Pentru fiecare dintre variabilele utilizate la sortare se poate stabili ordinea ascendentă sau descendentă. De notat că ordonarea bazei de date se face întâi în funcție de prima variabilă apoi în funcție de a doua în categoriile celei dintâi, etc.

Dacă am avea variabilele de sortare „Rezidență”, „Sex” și „Media anului trecut școlar”, întâi baza de date va fi împărțită în Rural și urban, apoi în categoriile de rezidență s-ar realiza ordonarea după apartenența de sex. În continuare, în fiecare categorie definită simultan după rezidență și sex se va face o ordonare după reușita școlară din anul precedent.

Graficul 6.37 Data Sort cases in SPSS

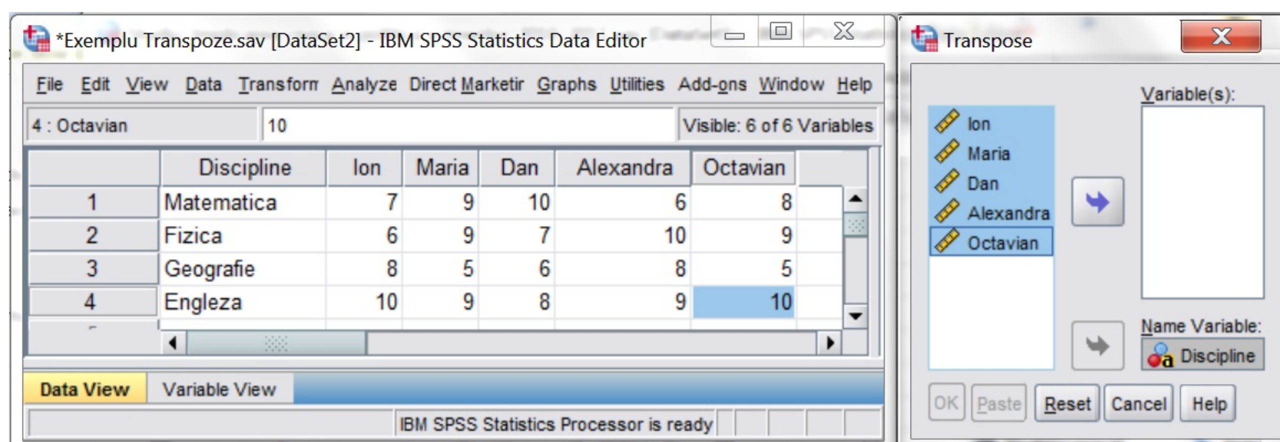


Opțiunea „Save file with sorted data” permite salvarea bazei de date în forma în care a fost ordonată (în funcție de variabilele indicate). Se poate salva ordinea inițială a variabilelor într-un atribut personalizat al variabilelor (vizibil în „Variable view”). Pentru fiecare variabilă, valoarea atributului este o valoare întregă care indică poziția sa înainte de sortare; astfel, prin sortarea variabilelor în funcție de acel atribut personalizat se poate restabili ordinea inițială a variabilelor (IBM SPSS Statistics 23 Brief Guide).

6.7 Variantele DATA, Transpoze, Merge files

Altă variantă a meniul „Data” este „Transpoze” care permite înlocuirea liniilor cu coloanele în baza de date. Prin acestă procedură cazurile devin variabile și variabile devin cazuri (Introduction to SPSS 19.0).

Graficul 6.38 Data Transpoze in SPSS



Să presupunem că avem o bază de date în care pentru patru discipline sunt trecute notele care au fost obținute de către cinci elevi (în exemplul nostru, elevii apar ca variabile).

Dacă dorim ca cele patru discipline să apară ca variabile iar elevii să aibă statutul de cazuri în baza de date SPSS, vom da comanda „Data Transpoze” şi vom obține fereastra de dialog din dreapta graficului 6.38. Caracteristicile desemnând numele elevilor vor fi transferte în căsuța „Variables” iar variabila „Discipline”, care desemnează materiile în discuție, va fi transferată în „Name Variable” (va da numele noilor variabile care vor apărea în urma transpunerii).

Graficul 6.39 Data Transpoze 2 in SPSS

Exemplu Transpoze 2.sav [DataSet3] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 5 of 5 Variables

	CASE_LBL	Matematica	Fizica	Geografie	Engleza	var
1	Ion	7	6	8	10	
2	Maria	9	9	5	9	
3	Dan	10	7	6	8	
4	Alexandra	6	10	8	9	

Data View Variable View

IBM SPSS Statistics Processor is ready

În graficul 6.39 se poate vedea că disciplinele de învățământ au devenit variabile iar elevii, împreună cu notele luate, reprezintă cazuri. De notat că valorile lipsă definite de utilizator (user missing) sunt convertite în date lipsă ale sistemului din fişierul de date transpus (system missing). Pentru a păstra valorile inițial definite ca „user missing”, trebuie definite ca date valide în ferestra „Variable View” a bazei de date inițiale (Oficial IBM SPSS Statistics 21 Brief Guide).

Următorea opțiune a meniului **Data** este **Merge files, Add cases** şi are rolul de a permite sumarea bazelor de date care conțin aceleași variabile dar cazuri diferite. În graficul 6.40 sunt prezentate două baze de date conținând acelaşi tip de informații dar cu unele deosebiri între variabile (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 6.40 Data Merge files, Add cases in SPSS 1

*Exemplu merge files add cases 1.sav [...

6 : Nr_frati — Visible: 4 of 4 Variables

	varsta	Nr_frati	Sex	Judet
1	12	1	Masc	Alba
2	15	2	Fem	Bacau
3	17	0	Fem	Cluj
4	33	3	Masc	Covasna
5	56	1	Masc	Bihor

*Exemplu merge files add cases 2.sav [DataS...

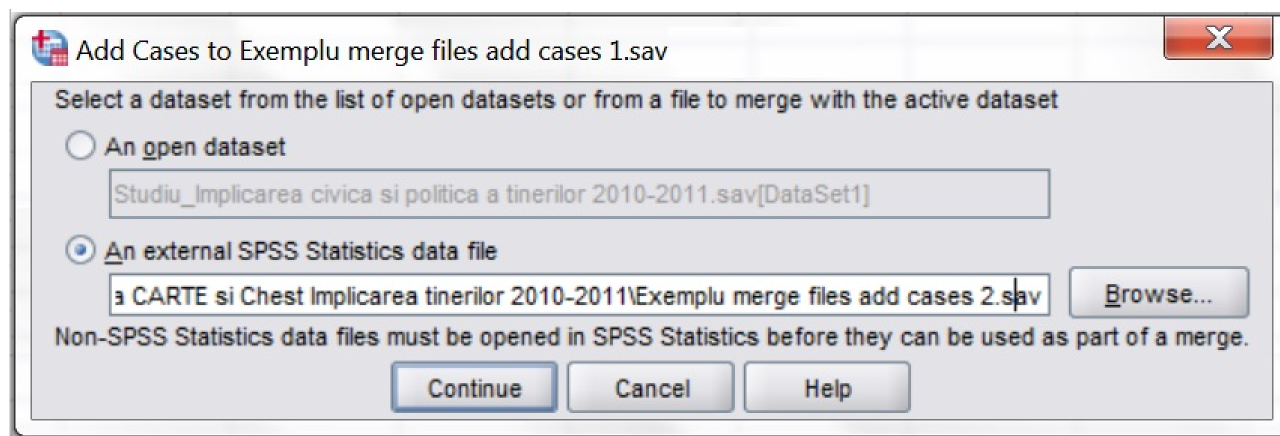
6 : Nr_frati — Visible: 4 of 4 Variables

	ani_impliniti	Nr_frati	Sex	Judet
1	22	0	Masc	Harghita
2	21	3	Fem	Sibiu
3	11	1	Masc	Prahova
4	13	2	Masc	Olt
5	14	0	Fem	Hunedoara

Pentru prima variabilă, informația are aceeași semnificație în ambele baze de date prezentate în graficul 6.40 dar numele variabilelor este diferit. Pentru variabila „Nr_frați", toate caractristicile sunt identice în cele două baze de date iar pentru caracteristica „Sex" diferă modul de definire al variabilei (este numerică în baza de date din stânga și „string" – șir de caractere – în cea din dreapta). Variabila „Județ" este definită „string" în ambele baze de date dar diferă numărul de caractere alocat („Width").

Pentru a lansa procedura de adăugare a cazurilor, în primul rând trebuie activată o bază de date (în cazul nostru, cea din dreapta graficului 6.40, numită „Exemplu merge files add cases 1"). Se dă apoi comanda „Data Merge files, Add cases" și se obține fereastra ilustrată în graficul 6.41 (SPSS: Data Manipulation and Advanced Topics For Windows).

Graficul 6.41 Data Merge files, Add cases in SPSS 2



Se poate observa că avem două variante pentru a selecta fișierul din care se vor adăuga cazuri: „An open dataset", adică o bază de date care a fost dechisă deja, sau „An external SPSS Statistic data file" (o bază de date SPSS care nu e încă deschisă și a cărei poziție se poate indica cu ajutorul căsuței aflată în fața butonului „Browse...."

În partea de jos a ferestrei apare un avertisment: bazele de date care nu sunt de tipul celor din programul SPSS (cu extensia .sav) trebuie întâi deschise cu SPSS pentru a putea urma procedura de unire a fișierelor prin adăugare de cazuri.

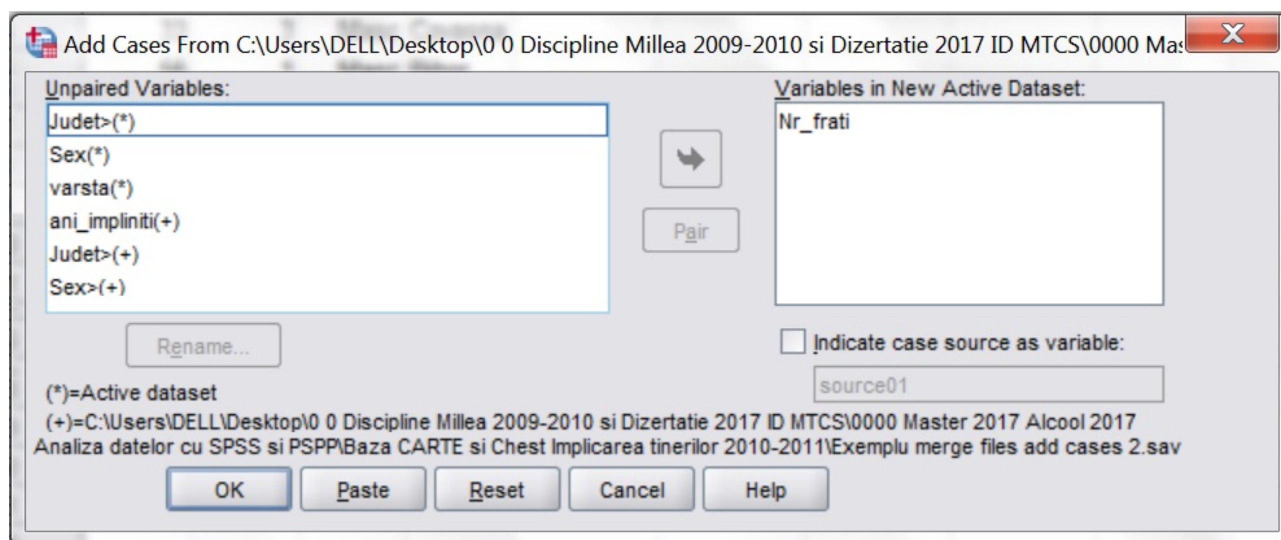
În graficul 6.41 se poate vedea că fișierul selectat pentru a adăuga noi cazuri este unul extern („An external SPSS Statistic data file") și poartă numele de „Exemplu merge files add cases 2". Dând comanda „Continue" vom obține o nouă fereastră de dialog (a se vedea graficul 6.42).

În partea din drepta a graficului 6.42 se află variabilele care au fost cuprinse de la început în noua bază de date sumativă deoarece sunt identic definite atât în fișierul activ cât și în cel extern. Menționăm faptul că, în stânga graficului 6.42, apare lista variabilelor care nu sunt definite în mod identic (Unpaired Variables). În respectiva listă, variabilele cu semnul „*" provin din fișierul activ iar cele cu semnul „+" aparțin fișierului extern.

Dacă știm că în cele două baze de date variabilele „ani_impliniti" și „varsta" au semnficație similară, putem să le selectăm pe ambele, ținând butonul „Ctrl" al tastaturii apăsat după care dăm click pe butonul „Pair". Rezultatul va fi că în viitoarea bază de date va apărea variabila „varsta" (nume preluat din fișierul activ) conținând atât informațiile din fișierul activ cât și din cel extern (SPSS: Data Manipulation and Advanced Topics For Windows).

Dacă dorim ca în noua bază de date să apară și variabilele care nu sunt pereche acestea vor trebui selectate din „Unpaired Variables” și trimise în „Variables in new Active Dataset”. Având în vedere că avem variabile cu același nume, va fi nevoie să schimbăm numele celei de a doua pe care o introducem în noua bază de date. Notăm faptul că variabilele „Județ”, deși au aceeași semnificație și sunt ambele „string”, nu pot fi trnsformate într-o singură variabilă deoarece au număr diferit de caractere (SPSS: Data Manipulation and Advanced Topics For Windows).

Graficul 6.42 Data Merge files, Add cases in SPSS 3



În graficul 6.43 e ilustrată baza de date care rezultă adăugând noi cazuri. Primele 2 variabile se referă la cazurile reunite („varsta” și „nr_frati”) pe când „Sex” și „Judet” vizează prima bază de date. Variabilele (redenumite) „sex2” și „Județ2” conțin informții doar despre fișierul extern.

De notat faptul că se poate alege prin selecție sau eliminare, care variabile din setul activ și din cel extern să facă parte din noua bază de date (SPSS: User Guide, Indian Council of Social Science Research).

Graficul 6.43 Data Merge files, Add cases in SPSS 4

*Exemplu merge files add cases 1.sav [DataSet5] - IBM SPSS S...

	varsta	Nr_frati	Sex	Judet	Sex2	Judet2
1	12	1	Masc	Alba		
2	15	2	Fem	Bacau		
3	17	0	Fem	Cluj		
4	33	3	Masc	Covasna		
5	56	1	Masc	Bihor		
6	22	0	.		Masc	Harghita
7	21	3	.		Fem	Sibiu
8	11	1	.		Masc	Prahova
9	13	2	.		Masc	Olt
10	14	0	.		Fem	Hunedoara

Cu ajutorul meniului Data pot fi unite baze de date și prin adăugarea de noi variabile fișierului activ folosind comenzile „Data, Merge files, Add Variables”. Procedura este utilă atunci când despre aceleași unități statistice s-au cules date la momente de timp distincte sau din surse diferite (informațiile despre elevi pot proveni de la ei înșiși, de la părinți, de la profesori, etc.) Pentru ca operațiunea să reușească este esențial ca baza de date activă și cea externă să fie sortate ascendent după una sau mai multe variabile identic definite în cele două baze de date și care să permită o ordonare perfectă în ambele cazuri. Altfel spus, e necesar ca cele două baze de date să conțină aceleași unități statistice / cazuri (SPSS: Data Manipulation and Advanced Topics For Windows).

Graficul 6.44 Data Merge files, Add Variables in SPSS 1

*Exemplu merge files add variables 1.sav [...

6 : Tot_gosp | 3 | Visible: 4 of 4 Variables

	Nr_chest	med5gen	Tot_gosp	Tot_frati
1	1	8-8.49	3	0
2	2	8.50-8.99	4	1
3	3	8-8.49	3	0
4	4	9-9.49	5	2
5	5	9.50-10	2	0
6	6	9.50-10	3	1

Data View | Variable View

*Exemplu merge files add variables 2.sav...

6 : Sex | 2 | Visible: 5 of 5 Variables

	Nr_chest	clasa	Varsta	Sex	Tot_gosp
1	1	5	13	1	3
2	2	7	14	2	4
3	3	9	15	2	5
4	4	10	18	1	3
5	5	12	20	1	4
6	7	12	19	2	6

Data View | Variable View

În graficul 6.44 se poate vedea că am ales două baze de date care au comună variabila Nr_chest, aceeași variabilă în funcție de care au fost ordonate ascendent ambele fișiere. Reamintim că sortarea fișierelor se putea face și după mai multe variabile pentru a obține o ordonare perfectă.

Alegând fișierul activ „Exemplu merge files add variables 1” și dând comanda „Data Merge files, Add Variables” obținem fereastra de dialog din graficul 6.45 în care putem alege o bază de date deja deschisă („Exemplu merge files add variables 2”) sau putem opta pentru o bază de date externă. În cazul nostru vom opta pentru „An open dataset” și apoi vom da comanda „Continue” (SPSS: User Guide, Indian Council of Social Science Research).

Graficul 6.45 Data Merge files, Add Variables in SPSS 2

Add Variables to Exemplu merge files add variables 1.sav[DataSet1]

Select a dataset from the list of open datasets or from a file to merge with the active dataset

An open dataset

Exemplu merge files add variables 2.sav[DataSet2]

An external SPSS Statistics data file

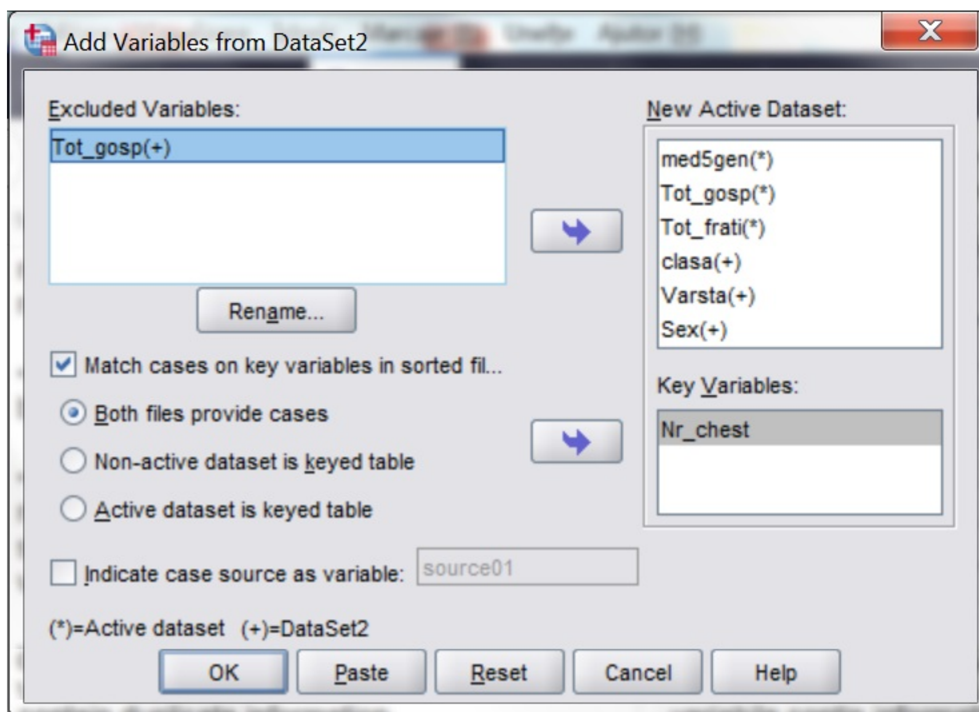
Browse...

Non-SPSS Statistics data files must be opened in SPSS Statistics before they can be used as part of a merge.

Continue | Cancel | Help

Am obținut, în continuare, fereastra din graficul 6.46, în care variabila „Nr_chest" a fost definita drept „Key Variable" adică cea care asigură potrivirea cazurilor din cele două fișiere. Se pot folosi una sau mai multe variabile cheie dar este necesar ca, în prealabil, bazele de date să fie sortate ascendent în funcție de fiecare dintre respectivele variabile (SPSS: Data Manipulation and Advanced Topics For Windows).

Graficul 6.46 Data Merge files, Add Variables in SPSS 3



De notat că o variabilă din a doua bază de date, simbolizată cu (+) al cărei nume e identic cu al unei variabile din prima bază de date, simbolizată cu (*), va fi poziționată în căsuța „Excluded Variables" și nu va fi cuprinsă în noul fișier decât dacă va fi redenumită („Rename") și va fi introdusă în căsuța „New Active Dataset:". În cazul nostru variabila „Tot_gosp(+)" ar putea avea semnificația de venit total al gospodăriei spre deosebire de „Tot_gosp(*)" care ar putea desemna numărul total al membrilor gospodăriei. În consecință, ar fi justificat ca ambele variabile să se regăsească în baza de date finală.

Dacă variabila / variabilele în funcție de care se face potrivirea cazurilor nu permit corespondența perfectă a acestora și am ales ca ambele fișiere să furnizeze cazuri („Both files provide cases"), variabilele care nu au corspondent vor cuprinde informații doar din baza de date din care provin. În graficul 6.47 e ilustrată baza de date rezultată din unirea celor două prezentate în graficul 6.44.

Se observă că variabila „Nr_chest", folosită pentru sortarea ascendentă a bazelor de date și pentru potrivirea cazurilor, a permis construirea unui fișier cu date complete pentru primele 5 cazuri dar pentru cazurile având numerele 6 și 7 datele au fost preluate doar din fișierele din care au provenit acestea (SPSS: Data Manipulation and Advanced Topics For Windows).

Notăm faptul că în unele situații este necesar ca una dintre bazele de date să fie considerată „Keyed table" – fișier de căutare – care să poată furniza valorile unui caz propriu pentru mai multe cazuri ale celuilalt fișier.

Graficul 6.47 Data Merge files, Add Variables in SPSS 4

*Exemplu merge files add variables 1.sav [DataSet3] - IBM SPSS Statistics Data Editor

	Nr_chest	med5gen	Tot_gosp	Tot_frati	clasa	Varsta	Sex	Ven_gosp	var
1	1	8-8.49	3	0	5	13	Ma...	3	
2	2	8.50-8.99	4	1	7	14	Fe...	4	
3	3	8-8.49	3	0	9	15	Fe...	5	
4	4	9-9.49	5	2	10	18	Ma...	3	
5	5	9.50-10	2	0	12	20	Ma...	4	
6	6	9.50-10	3	1	.	.	.	.	
7	7	.	.	.	12	19	Fe...	6	

Astfel de situații pot interveni, de exemplu, dacă un fișier conține date despre gospodărie (număr de camere, suprafață locuibilă, rezidență, diferite dotări, etc.) iar a doua bază de date conține informații despre fiecare membru de familie în parte (vârstă, studii, statut ocupațional, venit personal, etc.). „Key Variable" va avea valori unice pentru baza de date conținând informațiile despre gospodărie (definita ca fiind „Keyed table") dar va avea valori multiple pentru membrii aceleiași gospodării (a se vedea graficul 6.48).

Graficul 6.48 Data Merge files, Add Variables in SPSS 5

*Exemplu merge files add variables 3.sav [DataSet2...

	Gospodaria	Rezidenta	Nr_camere	Suprafata
1	1	Urban	2	50
2	2	Urban	3	100
3	3	Rural	5	200
4				
5				
6				
7				
8				
9				

*Exemplu merge files add variables 4.sav [...

	Gospodaria	Varsta	Venit	Nivel_Studii
1	1	23	1,800	scazut
2	1	55	3,400	ridicat
3	1	49	2,000	mediu
4	2	17	0	scazut
5	2	56	2,700	mediu
6	2	60	4,000	ridicat
7	3	22	1,500	scazut
8	3	54	4,500	ridicat
9	3	60	3,200	mediu

Variabila „Gospodăria" e cea care face legătura între cele două baze de date („Key Variable") iar fișierul care furnizează datele cazurilor proprii înspre mai multe cazuri ale celuilalt fișier este ilustrat în stânga graficului 6.48. În baza de date sumativă, celor din aceeași gospodărie (dreapta graficului 6.48) li se vor atașa valori identice pentru rezidența, numărul de camere și suprafața locuinței (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

După comanda „Data Merge files, Add Variables”, atunci când se ajunge la fereastra de dialog ilustrată în graficul 6.46, se va alege „Match cases on key variables in sorted file” (cu variabila de potrivire „Gospodăria” – „Key Variables”) și apoi se alege „Active dataset is keyed table” (fișierul activ este cel din stânga graficului 6.48 care conține variabilele „Rezidenta”, „Nr_camere” și „Suprafata”). În final va rezulta baza de date din graficul 6.49 (SPSS: Data Manipulation and Advanced Topics For Windows).

Graficul 6.49 Data Merge files Add Variables 6

*Exemplu merge files add variables 4.sav [DataSet2] - IBM SPSS Statistics Data Edit...

	Gospodaria	Varsta	Venit	Nivel_Studii	Rezidenta	Nr_camere	Suprafata
1	1	23	1,800	scazut	Urban	2	50
2	1	55	3,400	ridicat	Urban	2	50
3	1	49	2,000	mediu	Urban	2	50
4	2	17	0	scazut	Urban	3	100
5	2	56	2,700	mediu	Urban	3	100
6	2	60	4,000	ridicat	Urban	3	100
7	3	22	1,500	scazut	Rural	5	200
8	3	54	4,500	ridicat	Rural	5	200
9	3	60	3,200	mediu	Rural	5	200

Menționăm faptul că și în acest caz, pentru o combinare corectă a celor două baze de date, trebuia realizată o sortare ascendentă prealabilă a fiecăreia dintre acestea în funcție de variabila sau devariabilele care au fost definite drept „Key Variables”.

O altă opțiune a meniului „Data” este „Data Restructure” care are ca scop restructurarea unor variabile / cazuri din baza de date astfel încât să devină potrivite pentru analiza statistică. Restructurarea poate avea loc în sensul transformării variabilelor selectate în cazuri, a cazurilor selectate în variabile sau a transpunerii tuturor datelor: cazurile devin variabile și invers (SPSS: User Guide, Indian Council of Social Science Research).

6.8 Variantele DATA, Restructure selected variables into cases, Data Restructure selected Cases into Variables

Pentru a restructura datele există un îndrumător („Wizard”) lansat prin comanda „Data Restructure” care, inițial, permite selectarea restructurării dorite oferind informații scrise și ilustrate grafic pentru fiecare variantă.

„Restructure selected variables into cases” se utilizeză atunci când fiecare caz din din baza de date curentă are valori pentru anumite variabile care trebuie rearanjate în grupuri de cazuri aflate în relație în noua bază de date.

Să presupunem că într-o bază de date au fost introduse, ca variabile, performanțele înregistrate de 4 sportivi în anii 2014-2018, la două probe (100 și 800 m), variabilele constând în timpii întegistrați. Noi ne propunem să restructurăm cele două grupuri de variabile ce conțin informații despre aceeași probă sportivă și să le transpunem în 2 variabile în noua bază de date (SPSS: User Guide, Indian Council of Social Science Research).

În graficul 6.50 este prezentată baza de date pe care ne propunem s-o restructurăm, conținând date depre caracteristicile și performanțele a 4 sportivi pe parcursul a 4 ani pentru 2 probe sportive (100 respectiv 800m).

Graficul 6.50 Data Restructure selected variables into cases 1 in SPSS

Exemplu Restructure selected variables into cases 1 BIS.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

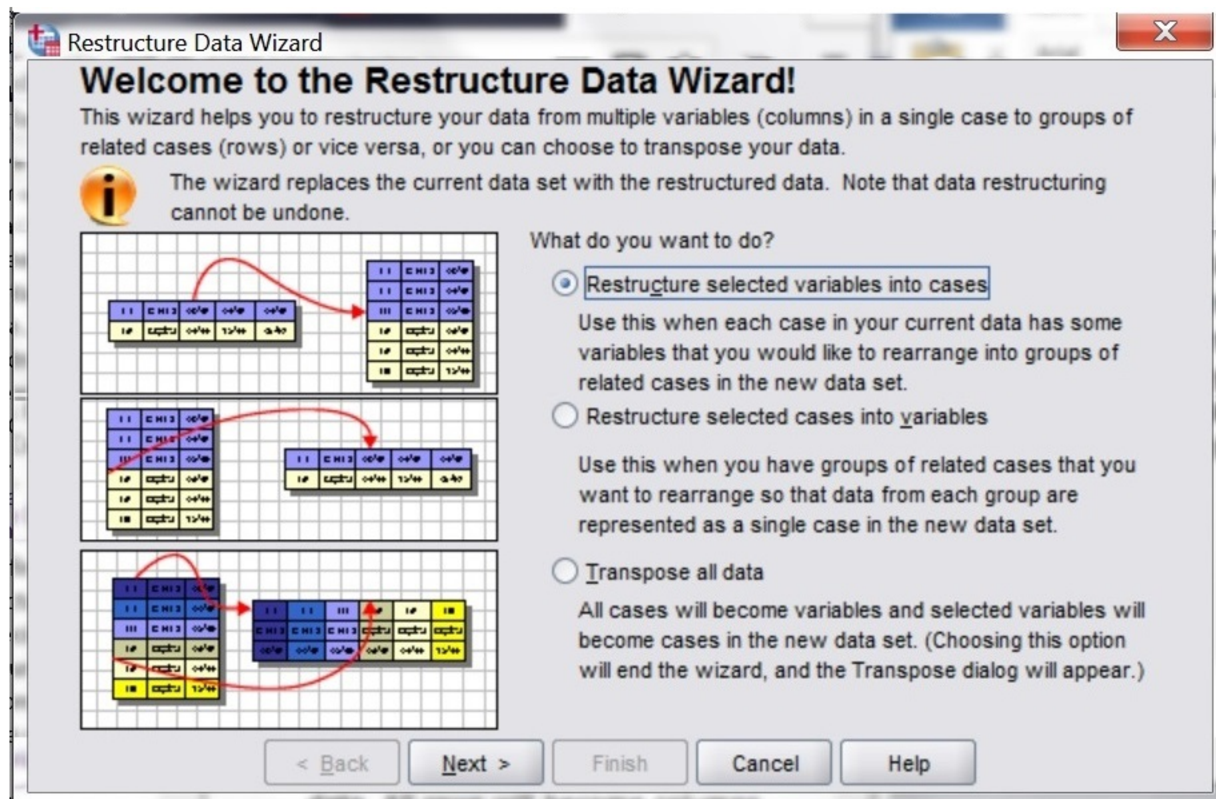
Visible: 11 of 11 Variables

	ID_SPORTIV	Varsta	Sex	An2015_100m	An2016_100m	An2017_100m	An2018_100m	An2015_800m	An2016_800m	An2017_800m	An2018_800m
1	Sportiv 1	19	Masc	10.01	9.89	10.01	9.95	1.75	1.69	1.73	1.69
2	Sportiv 2	24	Fem	9.94	9.88	9.99	9.89	1.73	1.69	1.77	1.71
3	Sportiv 3	23	Masc	9.99	9.86	9.89	9.96	1.72	1.71	1.73	1.72
4	Sportiv 4	39	Fem	10.20	9.99	9.95	9.88	1.67	1.68	1.69	1.68

Data View Variable View

Primul pas urmat în „Restructure Data Wizard” îl constituie selectarea tipului de transformare dorită, în cazul nostru „Restructure selected variables into cases” (graficul 6.51).

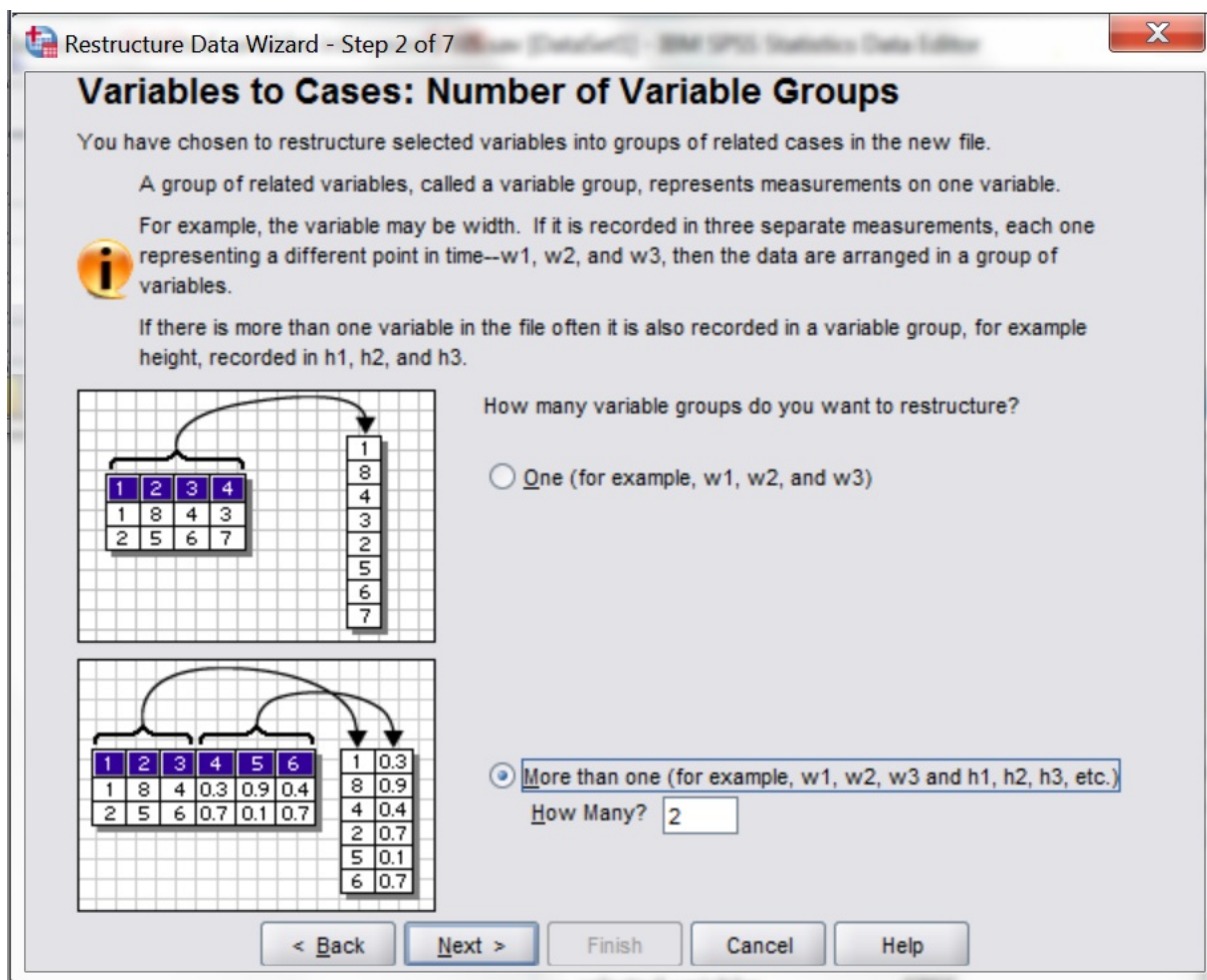
Graficul 6.51 Data Restructure selected variables into cases 2 in SPSS



In pasul următor se stabilește numărul de grupuri de variabile care se doresc a fi restructurate, respectiv transformate în cazuri în noua bază de date (SPSS: User Guide, Indian Council of Social Science Research).

În graficul 6.52 e ilustrat al doila pas al demersului de restructurare a grupurilor de variabile în grupuri de cazuri aflate în relație în noua bază de date. În cazul nostru, avem două grupuri de variabile aflate în relație, primul referindu-se la timpii înregistrați la proba de 100m în anii 2015-2018 și al doilea la timpii obținuți la proba de 800m. Ca atare, în graficul 6.52 se alege varianta „More that one...” pentru numărul de grupuri de variabile („How many variable groups do you want to restructure?”) și se specifică valoarea 2.

Graficul 6.52 Data Restructure selected variables into cases 3 in SPSS



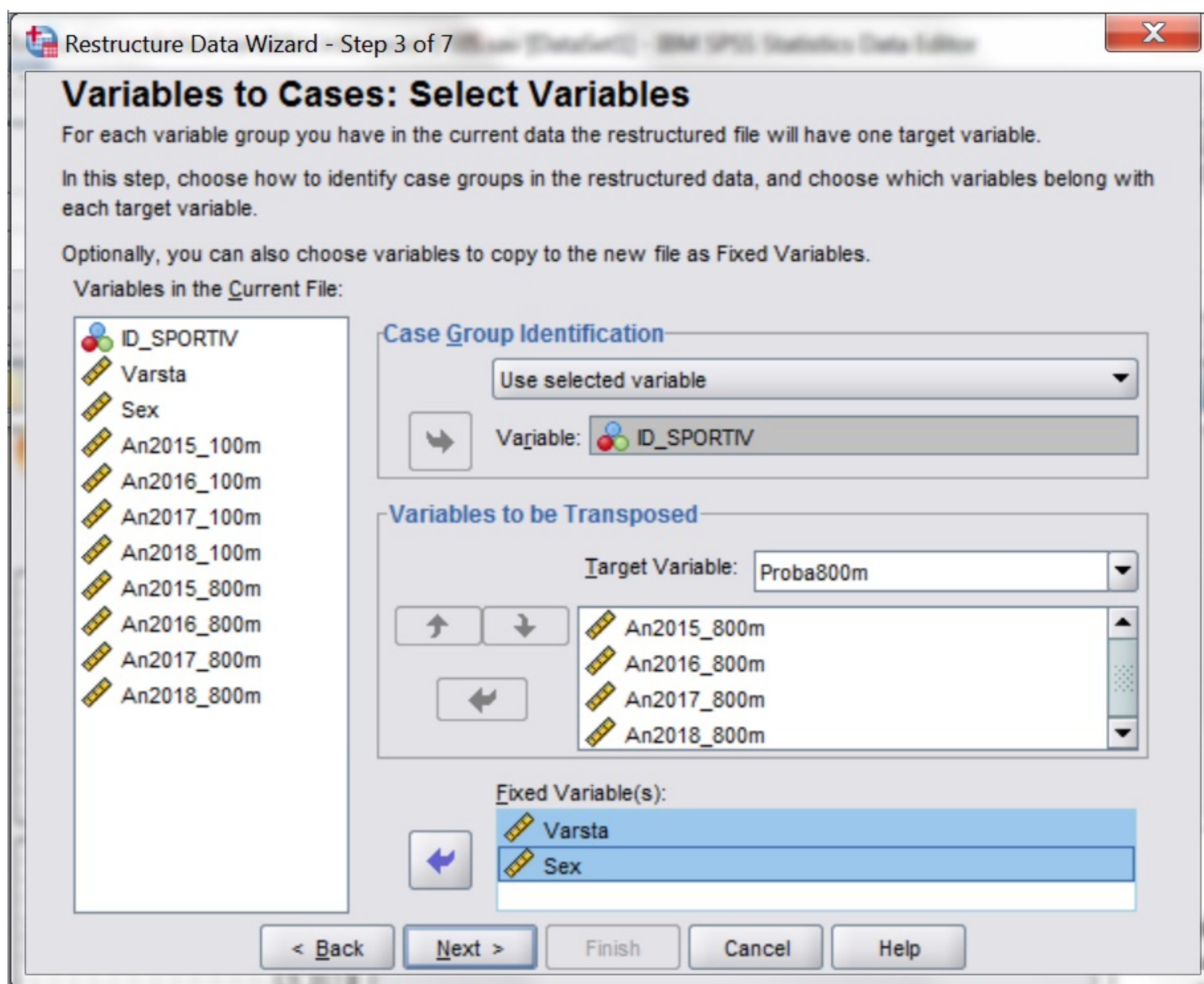
În graficul 6.53 este prezentat al treilea pas al demersului de rstructurare a bazei de date („Variables to cases”) astfel încât fiecărui grup de variabile îi va corespunde doar o variabilă („Target variable”).

Prima opțiune din fereastra de dialog „Case grup identification” permite ca în variabilele din baza de date restructurată să se găsească și un element de identificare semnificând cazurile din baza de date inițială. În cazul nostru, variabile este „ID_SPORTIV”, luînd valorile Sportiv1 - Sportiv4 pentru cei patru sportivi despre care există informații în baza de date inițală.

Următoarea căsuță, „Variables to be Transposed” permite să se facă legătura dintre fiecare grup de variabile și cea corespunzătoare: variabilele An2015_100m - An2018_100m au fost alocate variabilei Proba100m iar An2015_800m - An2018_800m au fost alocate variabilei Proba 800m (SPSS User Guide, Indian Council of Social Science Research).

Cu ajutorul opțiunii „Fixed variables" se pot copia variabile din vechea în noua bază de date asfel încât să se păstreze corespondența dintre cazurile inițiale și valorile luate pentru respectivele variabile. În cazul nostru variabilele vârstă și sex au fost cele transferate prin această modalitate.

Graficul 6.53 Data Restructure selected variables into cases 4 in SPSS



Pasul următor îl constituie crearea unei / unor variabile index menite să identifice grupurile de noi cazuri care au fost create pernind de la cazurile inițiale.

În exemplul nostru vom alege un singur index (SPSS User Guide, Indian Council of Social Science Research).

Fereastra următoare care apare în „Restructure Data Wuzard" se referă la caracteristicile indexului, care pot lua forma numerelor întregi în ordine crescătoare sau pot fi numele variabilelor. În cazul nostru este preferabilă prima variantă, în care 1, 2, 3 și 4 vor semnifica anii 2015, 2016, 2017, și 2018.

În graficul 6.54 e prezentată baza de date obținută în urma restructurării celei inițiale. Celor 4 variabile An2015_100m - An2018_100m le corespunde noua variabilă „proba100" iar celuilalt grup , An2015_800m - An2018_800m, îi corespunde variabila „proba 800".

Variabila ID_SPORTIV ne permite să știm la ce caz al bazei inițiale se referă variabilele proba100m și proba800m (Sportiv1-Sportiv4) iar variabila „Anul" ne informează despre momentul la care se referă înregistrarea IBM SPSS Statistics 20 Core System, User's Guide, 2011.

Graficul 6.54 Data Restructure selected variables into cases 5 in SPSS

*Exemplu Restructure selected variables into cases 1 BIS.sav [DataSet1] - IBM SPSS Statistics Data...

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	ID_SPORTIV	Varsta	Sex	Anul	proba100	proba800	var	var
1	Sportiv 1	19	Masc	1	10.01	1.75		
2	Sportiv 1	19	Masc	2	9.89	1.69		
3	Sportiv 1	19	Masc	3	10.01	1.73		
4	Sportiv 1	19	Masc	4	9.95	1.69		
5	Sportiv 2	24	Fem	1	9.94	1.73		
6	Sportiv 2	24	Fem	2	9.88	1.69		
7	Sportiv 2	24	Fem	3	9.99	1.77		
8	Sportiv 2	24	Fem	4	9.89	1.71		
9	Sportiv 3	23	Masc	1	9.99	1.72		
10	Sportiv 3	23	Masc	2	9.86	1.71		
11	Sportiv 3	23	Masc	3	9.89	1.73		
12	Sportiv 3	23	Masc	4	9.96	1.72		
13	Sportiv 4	39	Fem	1	10.20	1.67		
14	Sportiv 4	39	Fem	2	9.99	1.68		
15	Sportiv 4	39	Fem	3	9.95	1.69		
16	Sportiv 4	39	Fem	4	9.88	1.68		

Variabilele „Vârstă” și „Sex” au fost copiate așa încât să fie păstrată legătura cu cazurile inițiale din baza de date (Sportiv 1 – Sportiv 4).

Comanda „Data Restructure” are și opțiunea „Select Cases into Variables” care este utilizată atunci când există grupuri de linii aflate în relație în baza de date și se dorește ca ele să apară în grupuri de coloane (variabile) în baza de date restructurată. În graficul 6.55 e prezentată baza de date în care liniile sunt în relație prin intermediul momentului testării, operaționlizat prin variabila „Moment” (SPSS User Guide, Indian Council of Social Science Research).

Graficul 6.55 Data Restructure selected Cases into Variables 1 în SPSS

Exemplu Restructure selected cases into variables 1.sav [DataSet2] - IBM SPSS Statistics Data Editor

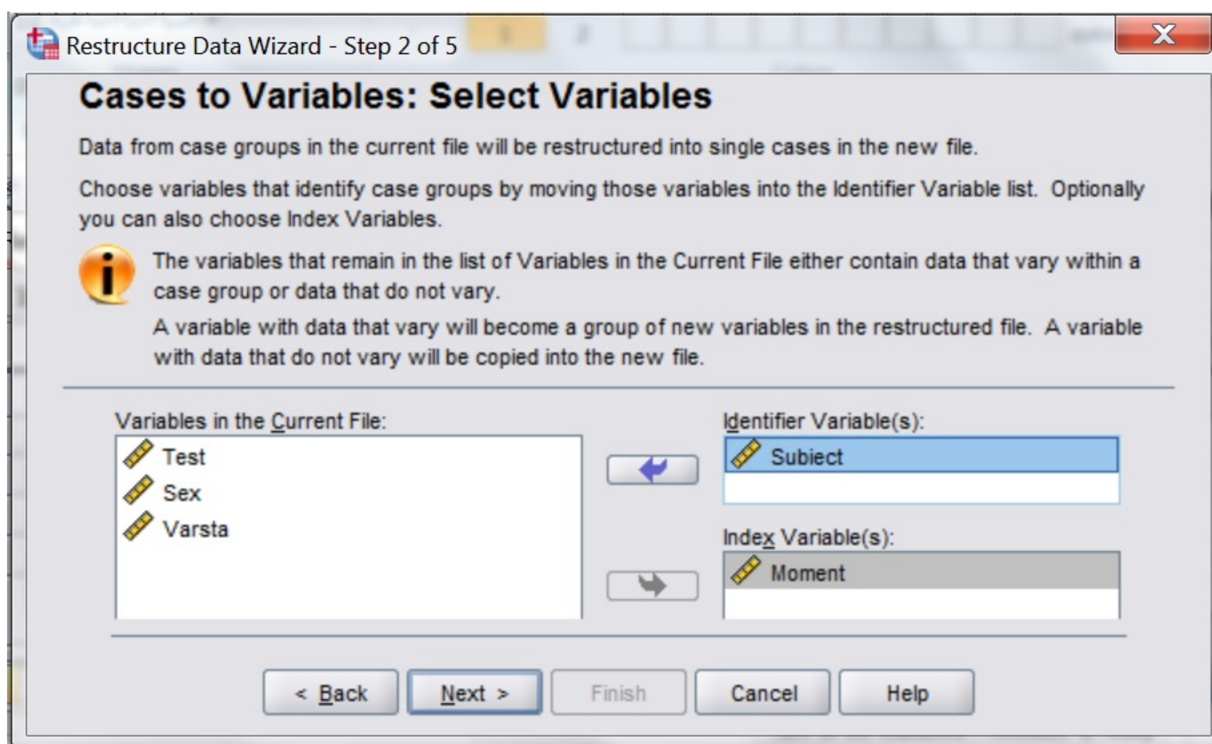
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Subiect	Test	Moment	Sex	Varsta	var
1	Subiect1	7	Inceput	Masculin	16	
2	Subiect1	9	Final	Masculin	17	
3	Subiect2	6	Inceput	Feminin	17	
4	Subiect2	8	Final	Feminin	18	
5	Subiect3	5	Inceput	Masculin	17	
6	Subiect3	7	Final	Masculin	17	
7	Subiect4	6	Inceput	Feminin	16	
8	Subiect4	9	Final	Feminin	17	

Data View Variable View

În baza de date există informații despre patru subiecți care au fost testați la două momente de timp (Început și Final) iar ca variabile adiționale au fost înregistrate apartenența de sex și vârsta la primul și la al doilea test. Prin restructurarea bazei de date se urmărește ca datele privitoare la cele două momente de timp sa apară în variabile distincte.

Graficul 6.56 Data Restructure selected Cases into Variables 2 în SPSS



Identificatorul pentru variabile a fost reprezentat de subiectul la care se referă datele iar indexul variabilelor a fost ales momentul de timp la care au fost culese acestea. Variabila „Moment a luat valoarea 1 pentru „Inceput" și 2 pentru final și a fost utilizată pentru a construi variabilele care și-au schimbat valorile între „Început" și „Sfârșit". Pe liniile bazei de date din graficul 6.57 avem cazurile care se referă la fiecare subiect, variabila nemodificată „Sex" și variabilele vizând cele două teste și vârstele la momentele de timp 1 și 2 (SPSS User Guide, Indian Council of Social Science Research).

Graficul 6.57 Data Restructure selected Cases into Variables 2 în SPSS

*Exemplu Restructure selected cases into variables 1.sav [DataSet3] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 6 of 6 Variables

	Subiect	Sex	Test.1	Test.2	Varsta.1	Varsta.2
1	Subiect1	Masculin	7	9	16	17
2	Subiect2	Feminin	6	8	17	18
3	Subiect3	Masculin	5	7	17	17
4	Subiect4	Feminin	6	9	16	17

Data View Variable View

Ultima opțiune pentru „Data Restructure” este „Transpoze all data” care e identică cu procedura „Data Transpoze” prezentată în graficele 6.38 şi 6.39.

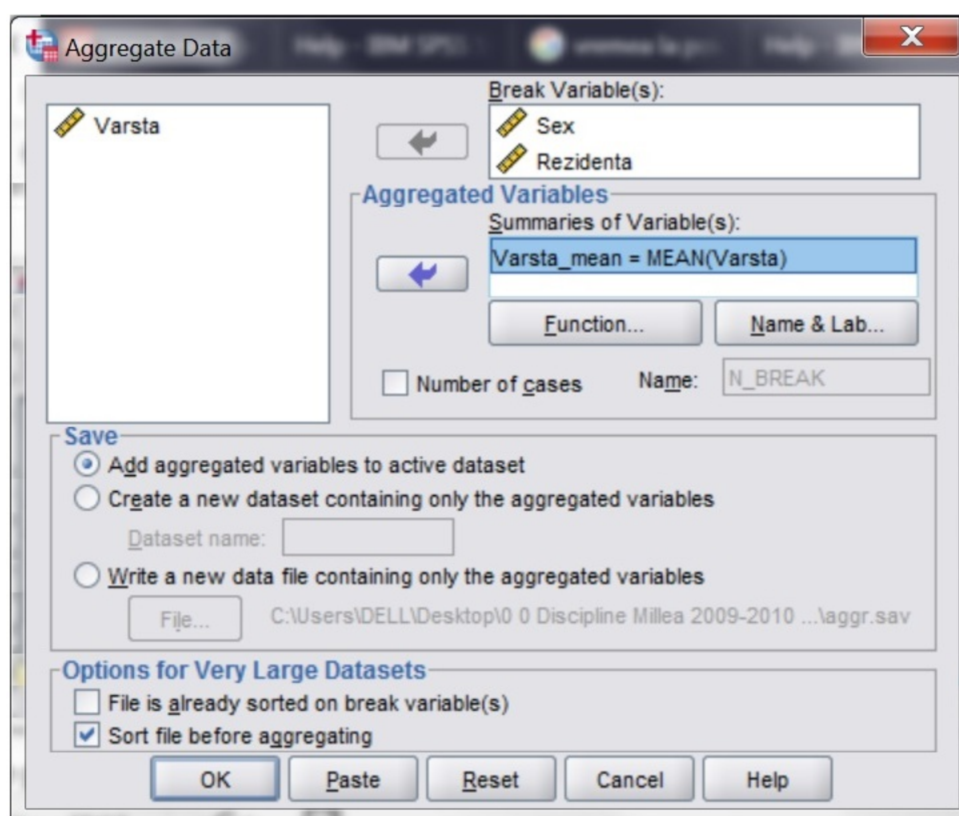
6.9 Variantele DATA, Agregate, Orthogonal Design, Split File, Copy Dataset, în SPSS Select Cases, Weight Cases

Prezentăm, succint, în continuare, comanda „Data Aggregate” având în vedere că este foarte asemănătoare cu cea existentă în PSPP (graficul 6.4).

Prin „Data Aggregate” se atribuie valori unice unor grupuri de cazuri ale unei variabile din baza de date activă şi se creează un nou fişier cu datele agregate sau se adaugă noi variabile (agregate) în baza de date existentă.

Cazurile sunt agregate pe baza valorilor uneia sau mai multor variabile de grupare („break variables”). Pentru fiecare combinație unică a valorilor variabilelor de grupare va fi genertă o singură valoare a variabilei agregate (IBM Knowledge Center, Home, SPSS Statistics 24.0).

Graficul 6.58 Data Aggregate 1 în SPSS



În graficul 6.58 e prezentată fereastra de dialog prin care putem stabili variabila vizată pentru agregare („Aggregated variables”: Vârsta) şi care sunt cele de grupare („Break variables”: rezidența şi sexul). Pentru variabila de agregare putem stabili modul în care se vor aloca valorile. În SPSS sunt disponibile mai multe funcții care pot fi aplicate asupra fiecărul grup de valori în care a fost împărțită variabila destinată agregări de către variabilele de grupare. În cazul nostru, în fiecare grup cu aceeaşi rezidență şi acelaşi sex se va calcula valoarea medie a vârstei (vârsta medie a bărbaților din urban, a femeilor din urban, a bărbaților din rural şi a femeilor din rural).

În graficul 6.59 este prezentată baza de date inițială - în partea din stânga - și baza SPSS rezultată în urma agregării - în partea din dreapta (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 6.59 Data Aggregate 2 în SPSS

*Exemplu Data Agregate SPSS.s...

	Varsta	Sex	Rezidenta
1	25	Feminin	Oras
2	66	Feminin	Oras
3	22	Feminin	Sat
4	29	Feminin	Sat
5	34	Masculin	Oras
6	56	Masculin	Oras
7	39	Masculin	Sat
8	43	Masculin	Sat

*Untitled3 [Exemplul2] - IBM SPSS St...

	Sex	Rezidenta	Varsta_mean
1	Feminin	Oras	45.50
2	Feminin	Sat	25.50
3	Masculin	Oras	45.00
4	Masculin	Sat	41.00
5			
6			
7			

Menționăm faptul că programulSPSS propune două modalități de salvare a variabilei agregate: adăugarea acesteia la baza de date inițială (atribuind aceleași valori în interiorul fiecărui grup) sau crearea unui nou fișier conținând doar variabila agregată și criteriile de agregare. A doua modalitate are două opțiuni, una limitându-se la crearea fișierului, cealaltă presupunând și salvarea acestuia (Introduction to SPSS 19.0).

Următoarea opțiune a meniului „Data" este „Orthogonal Design" care poate fi aplicată în cazul studierii preferințelor indivizilor pentru un anumit produs / serviciu, caracterizat prin mai mulți factori, fiecare dintre aceștia având mai multe niveluri de manifestare.

Primul pas într-o analiză de tip „conjoint" este să creezi combinații de factori, având diferite nivele, care vor fi prezentate subiecților evluatori sub formă de profiluri ale produselor / serviciilor oferite. Deoarece chiar și un număr mic de factori și câteva nivele ale fiecăruia conduc la un număr imens de profiluri ale produsului /serviciului, trebuie generat un subset reprezentativ al acestor profiluri, care este cunoscut sub numele de matrice ortogonală (IBM SPSS Statistics V22.0.0 documentation).

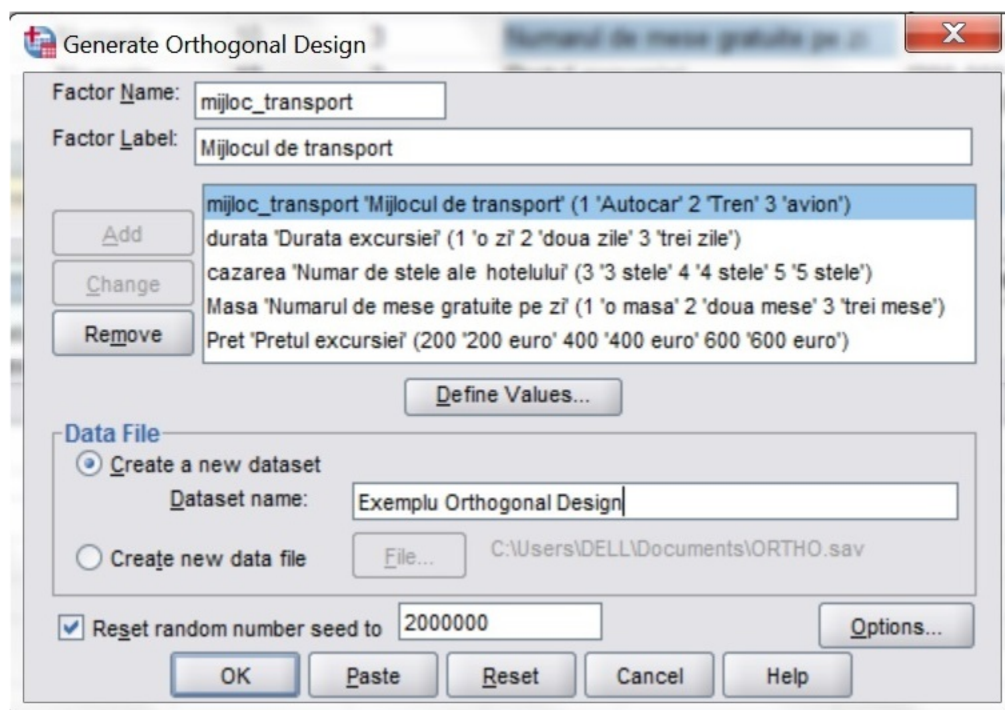
Procedura „Generate Orthogonal Design" creează o matrice ortogonală (design ortogonal) și stochează informațiile într-un fișier de date.

De notat că nu este necesară deschiderea unei baze de date care să conțină informații înainte de a executa procedura „Generate Orthogonal Design", se poate crea un nou fișier în care vor fi construite noile variabile (aici vor fi indicate numele, etichetele numelor precum și etichete ale valorilor variabilelor).

Lansând comanda „Data, Orthogonal Design, Generate" va apărea fereastra de dialog din graficul 6.60 care permite introducerea factorilor care sunt luați în considerare și a valorilor pe care acești le pot lua. În primul rînd, se va introduce numele factorului și apoi eticheta acestuia.

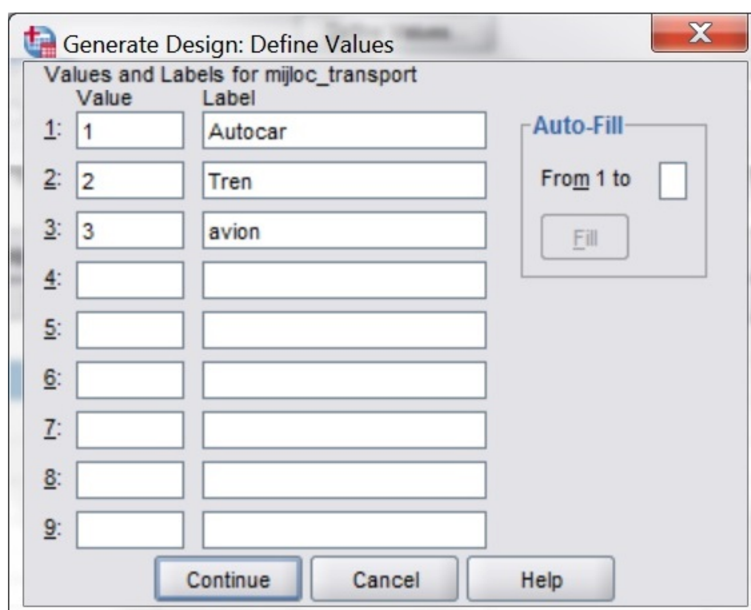
În cazul nostru, primul factor este „mijloc_transport" având eticheta „Mijlocul de transport" și valorile 1, 2 și 3 pentru Autocar, Tren și avion. Ceilalți factori care au fost introduși sunt „durata", „cazarea", „Masa" și „Pret" având etichetele variabilelor și ale valorilor acestora vizibile în graficul 6.60.

Graficul 6.60 Data, Orthogonal Design 1 în SPSS



Introducerea factorilor și a etichetelor lor se face în partea de sus a ferestrei de dialog (Graficul 6.60). Urmează comanda „Add". Apoi selectăm factorul nou introdus și dăm click pe „Define Values". Procedura apare în graficul 6.61 (IBM SPSS Statistics V22.0.0 documentation).

Graficul 6.61 Data, Orthogonal Design 2 în SPSS



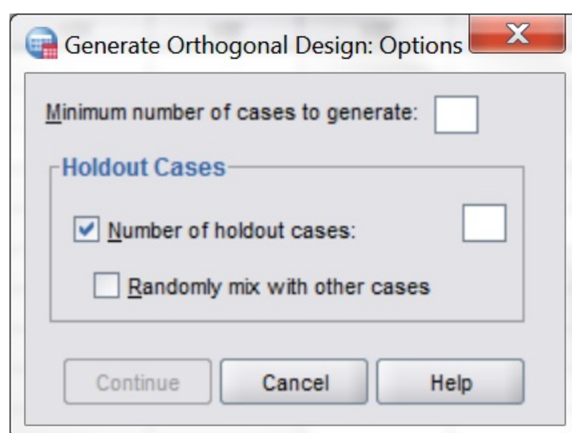
În fereastra de dialog din graficul 6.61 se vede că valorile variabilelor (factorilor) se introduc în partea stângă iar etichetele corespunzătoare se poziționează în partea dreaptă (după care se dă comanda „Continue").

De menționat că numele factorului pentru care se introduc datele e indicat în partea de sus a graficului. De asemenea, se poate vedea că numărul maxim de valori care poate fi introdus pentru un factor este 9.

În Fereastra de dialog din graficul 6.60 avem două opțiuni de salvare a fișierului care va fi generat de SPSS: „Create new dataset" care va genera setul de date cu rezultatele designului ortogonal fără a stabili folderul de destinație pentru fișier, respectiv „Create a new data file", care presupune alocarea unei destinații și a unui nume pentru o bază de date SPSS cu extensia .sav (IBM SPSS Statistics V22.0.0 documentation)

Tot în graficul 6.60 avem posibilitatea de a seta numărul aleator de selecții („Reset random number seeds to"). Generarea unui design ortogonal necesită un set de numere aleatoare. De reținut că dacă se dorește duplicarea unui design trebuie setată valoarea numărului de selecții înainte de a-l genera și trebuie resetată la aceeași valoare de fiecare dată când se reia procedura. În exemplul nostru numărul ales a fost 2000.000 (maximul e de 2.000.000.000.)

Graficul 6.62 Data, Orthogonal Design 3 în SPSS



În fereastra prezentată în graficul 6.60 există un buton „Options" (dreapta jos) care permite, în primul rând, stabilirea numărului minim de cazuri generate (Minimum number of cases to generate). Este important să reținem faptul că, implicit, SPSS generează automat numărul minim de cazuri dintr-o matrice ortogonală (numărul cel mai mic de cazuri care trebuie analizate de evaluatori pentru a se putea estima utilitățile: factorii și caracteristicilor acestora). În unele cazuri e utilă, totuși, specifcarea numărul minim de cazuri generate (de exemplu, în situația în care numărul implicit de cazuri este prea mic pentru ca analiza combinațiilor de factori și de valori ale acestora să fie utilă și / sau relevntă).

Opțiunea „Holdout Cases" se referă la cazurile care sunt supuse evaluării subiecților dar nu au fost generate prin designul ortogonal. În căsuța „Number of hold cases" se poate specifica orice număr întreg pozitiv inferior numărului total de cazuri care pot fi formate din toate combinațiile posibile de niveluri de factori. De notat că aceste cazuri sunt generate de un alt plan aleator, nu de planul experimental și nu generează cazuri identice cu cel experimental (IBM SPSS Statistics V22.0.0 documentation).

Selectând opțiunea „Randomly mix with other cases”, cazurile generate prin designul ortogonal vor fi amestecate cu celelalte, geneate de un alt plan aleator. Dacă butonul nu este selectat, cazurile vor fi prezentate separat.

Următorul pas îl constituie comanda „Ok” din fereastra de dialog prezentată în graficul 6.60. În exemplul nostru n-am selectat un număr minim de cazuri care să fie generate de designul ortogonal și n-am solicitat generarea unor cazuri suplimentare. În graficul 6.63 se vede că SPSS a generat 16 cazuri prin designul ortogonal (IBM Knowledge Center, Home, SPSS Statistics 24.0.0.)

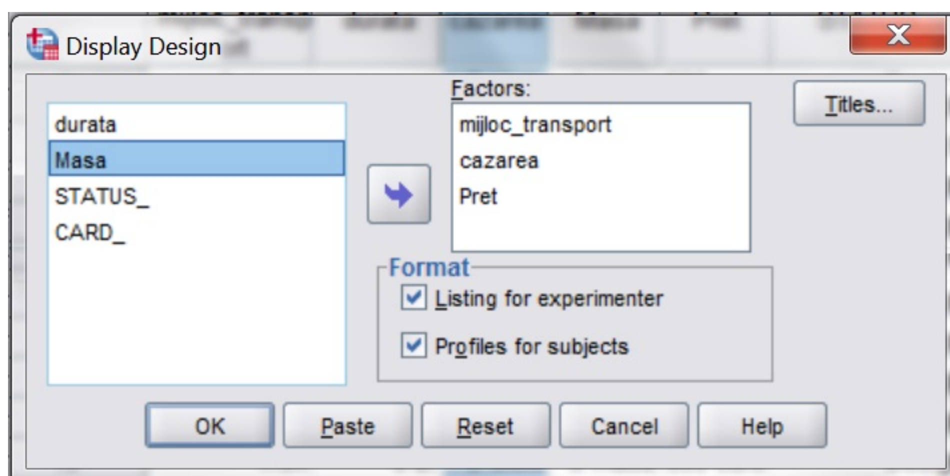
Graficul 6.63 Data, Orthogonal Design 4 în SPSS

	mijloc_transport	durata	cazare	Masa	Pret	STATUS_	CARD_	var
1	Autocar	o zi	5 stele	doua mese	600 euro	Design	1	
2	Autocar	doua zile	3 stele	trei mese	200 euro	Design	2	
3	Tren	trei zile	5 stele	o masa	200 euro	Design	3	
4	Autocar	o zi	5 stele	trei mese	400 euro	Design	4	
5	Autocar	o zi	3 stele	o masa	200 euro	Design	5	
6	avion	doua zile	5 stele	o masa	200 euro	Design	6	
7	Autocar	o zi	3 stele	o masa	200 euro	Design	7	
8	avion	o zi	3 stele	o masa	400 euro	Design	8	
9	Autocar	trei zile	3 stele	doua mese	200 euro	Design	9	
10	Tren	doua zile	3 stele	doua mese	400 euro	Design	10	
11	Autocar	trei zile	4 stele	o masa	400 euro	Design	11	
12	Autocar	doua zile	4 stele	o masa	600 euro	Design	12	
13	Tren	o zi	3 stele	o masa	600 euro	Design	13	
14	avion	o zi	4 stele	doua mese	200 euro	Design	14	
15	avion	trei zile	3 stele	trei mese	600 euro	Design	15	
16	Tren	o zi	4 stele	trei mese	200 euro	Design	16	

În consecință, în cadrul unei cercetări, subiecților li s-ar cere să compare (ierarhizeze) 16 variante de combinare a factorilor și valorilor lor.

Comanda „Data, Orthogonal Desing” are și opțiunea „Display” care permite printarea unui design experimental sub formă de draft sau ca profile care pot fi prezentate subiecților într-un studiu. În graficul 6.64 e prezentată fereastra obținută prin comenzile „Data”, Orthogonal Design”, Display (IBM SPSS Statistics V22.0.0 documentation).

Graficul 6.64 Data, Orthogonal Design Display în SPSS



În primul rând, din căsuța din partea stângă a graficului 6.64 trebuie deplasați în partea dreaptă aacestuia (căsuța „Factors”) unul sau mai mulți factori care vor fi prezentați în fișierul de ieșire (de tip output). În continuare, trebuie selectat un format (din căsuța „Format”) pentru prezentarea profilelor în fișierul output. Se poate opta pentru „Listing for experimenter” ceea ce presupune afișarea unui format draft în care se face diferența dintre profilele rezultate din designul ortogonal și cele generate suplimentar (profilele simulate vor fi listate separat în funcțiie de cele două moduri de generare). Dacă alegem „Profiles for subjects” vor fi listate profile cu factori și valori ale acestora care pot fi prezentate subiecților fără a se face o diferență între cele generate de designul ortogonal și cele suplimentare. O ultimă opțiune, „Titles” permite definirea de „headers” și „footers” atât pentru formatul draft destinat cercetătorului cât și pentru profilele din output care vor fi supuse evaluării respondenților anchetei (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

În meniul „Data”, următoarea opțiune este „Copy Dataset” care permite unui fișier bază de date din SPSS să fie multiplicat (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 6.65 Data, Copy Dataset în SPSS

*Exemplu constructie baza de date_bis.sav [DataSet9] - IBM SP...

File Edit View Data Transfor Analyz Direct Marke Graph Utilitie Add-on Windov Help

1 : V4mama 4 Visible: 7 of 7 Variables

	nr_chest	V1	V2	V3	V4mama
1	1	9.42	1-1.99 ore	Importanta	postliceala
2	2	6.54	sub 15 min	Deloc	max. 8 clase
3	3	8.11	15-29 min	Asa si asa	liceu
4	4	8.47	30-59 min	Putin	postliceala
5	5	9.83	2-2.99 ore	Importanta	studii superioare
6	6	9.36	3 ore sau...	Foarte imp...	studii superioare
7	7	6.31	NR	NR	NR

Data View Variable View

IBM SPSS Statistics Processor is ready

*Untitled9 [DataSet12] - IBM SPSS Statistics Data Editor

File Edit View Data Transfor Analyz Direct Marke Graph Utilitie Add-on Windov Help

Visible: 7 of 7 Variables

	nr_chest	V1	V2	V3	V4mama
1	1	9.42	1-1.99 ore	Importanta	postliceala
2	2	6.54	sub 15 min	Deloc	max. 8 clase
3	3	8.11	15-29 min	Asa si asa	liceu
4	4	8.47	30-59 min	Putin	postliceala
5	5	9.83	2-2.99 ore	Importanta	studii superioare
6	6	9.36	3 ore sau...	Foarte imp...	studii superioare
7	7	6.31	NR	NR	NR

Data View Variable View

IBM SPSS Statistics Processor is ready

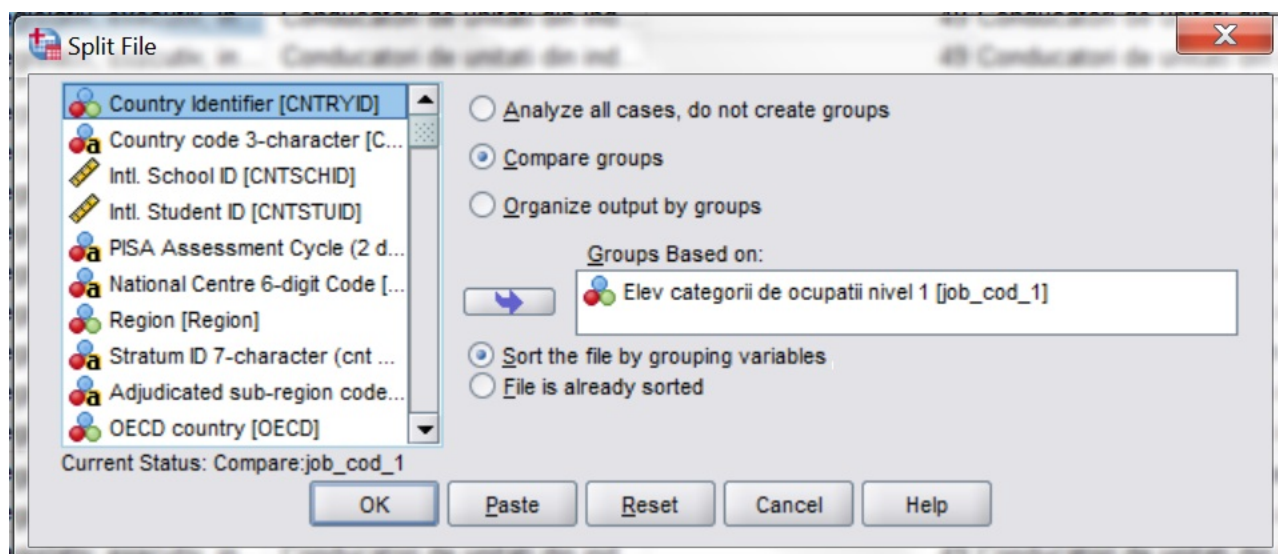
Dând comanda „Data”, „Copy Dataset”, se va genera un nou fișier, identic cu cel activ în programul SPSS pentru toate caracteristicile acestuia (nume variabile, mod de definire, valori, număr de cazuri), singura diferență constând, evident, în numele și destinația noii baze de date care urmează a fi stabilite.

Meniul **Data** oferă și opțiunea **Split File** care permite segmentarea bazei de date în funcție de aparteneța cazurilor la categoriile uneia sau mai multor variabile (variabilele trebuie să fie categoriale pentru ca instrucțiunea „splite file” să aibă sens). Cât timp comanda „Data”, „Split File” rămâne activă pentru un set de caracteristici, toate analizele statistice se vor realiza separat în interiorul categoriior definite de respectivele caracteristici.

Menționăm faptul că ordinea variabilelor este importantă: întâi baza de date va fi împărțită în categoriile primeia dintre acestea apoi fiecare fiecare grup de cazuri delimitat de prima variabilă va fi segmentat în funcție de a doua. În mod analog, vor fi luate în considerare următoarele variabile (pot fi cel mult 8). Cazurile trebuie ordonate după valorile variabilelor de grupare și în aceeași ordine în care apar în lista variabilelor (Groups Based On:). Dacă fișierul de date nu este deja sortat, trebie selectat „Sort the file by grouping variables” (SPSS Data Manipulation and Advanced Topics For Windows).

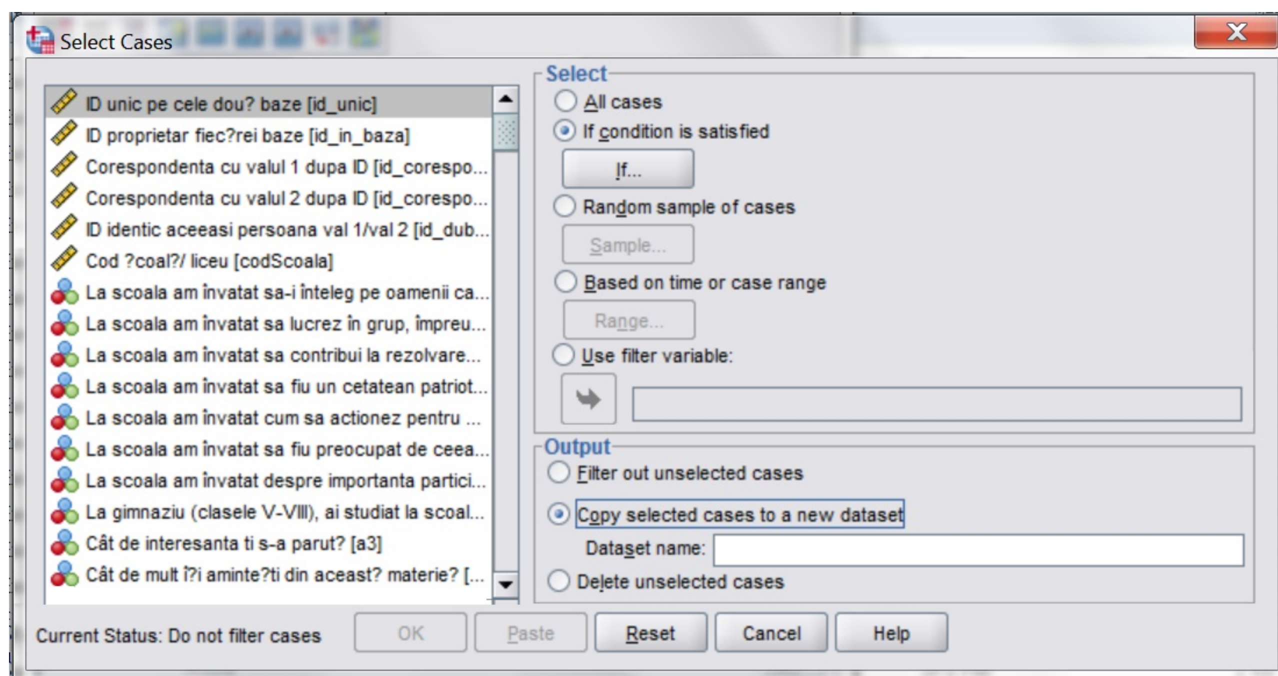
În momentul finalizării sesiunii de lucru, pentru a reveni la analiza întergii baze de date SPSS trebuie optat pentru „Analyze all cases, do not create groups" (sau se poate de click pe butonul „Reset" din partea de jos a ferestrei).

Graficul 6.66 Data, Split File în SPSS



Opțiunea „**Select Cases**" a meniului „Data" vizează selectarea anumitor cazuri din baza de date (conform unor criterii prestabilite) urmând ca analiza datelor să se facă doar asupra acestora.

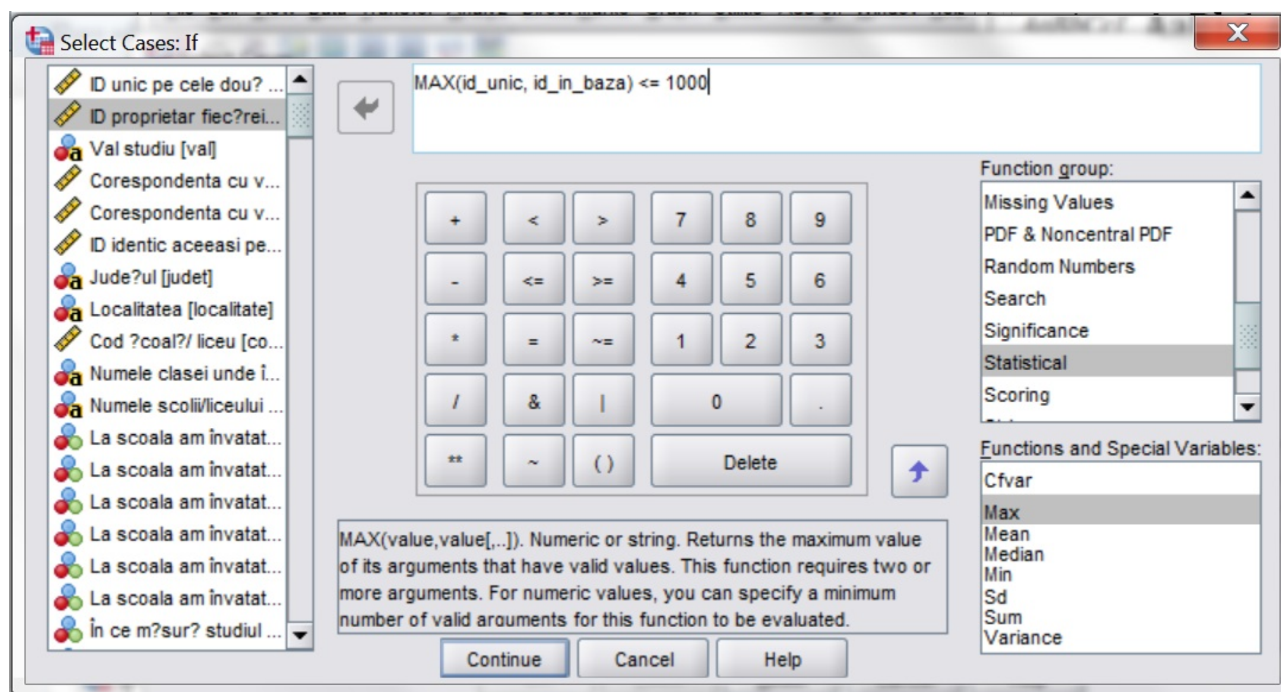
Graficul 6.67 Data, Select Cases 1 în SPSS



Prima modalitate de selectare a cazurilor o constituie formularea unei condiții logice („If condition is satisfied") care va conduce la selectarea cazurilor pentru care respectiva condiție este îndeplinită. În graficul 6.68 apare fereastra de dialog ce permite stabilirea condiției logice pentru selecție (SPSS Survival manual A step by step guide to data analysis).

În căsuța de sus a ferestrei din graficul 6.68 se pot introduce expresii simple, cum ar fi val = 1, care ar conduce la selectarea tuturor cazurilor pentru care variabila „Val studiu" este egală cu 1. De asemenea, se pot folosi funcțiile și grupurile de funcții disponibile în programul SPSS („Function group" cu funcțiile și variabilele speciale care le corespund, având titulatura „Functions and Special Variables").

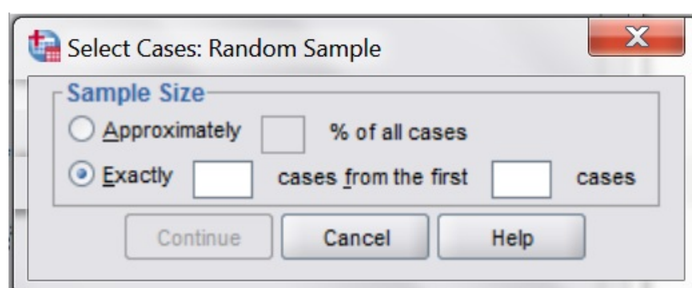
Graficul 6.68 Data, Select Cases If condition is satisfied în SPSS



În exemplul nostru, condiția logică pe care am formulat-o este ca valoarea maximă dintre variabilele „id_unic" și „id_in_baza" să fie mai mică sau egală cu 1000 (partea de sus a graficului 6.68). Am folosit grupul de funcții „Statistical" și funcția componentă „Max", care este descrisă în căsuța din partea de jos a ferestrei (Argumentele funcției, două sau mai multe, pot fi sau numerice sau string. Funcția returnează valoarea maximă dintre argumentele valide). Funcția poate fi selectată și deplasată în căsuța de sus a ferestrei iar variabilele vizate pot fi selectate din lista afaltă în stânga graficului 6.68 și poziționate în formulă (SPSS for Windows Version 19.0 A Basic Tutorial).

A doua modalitate de selectare a cazurilor presupune extragerea de eșantioane aleatoare din baza de date. În graficul 6.69 se vede că prima variantă constă în selectarea aproximativă a unui procent din totalul cazurilor (Oficial IBM SPSS Statistics 21 Brief Guide).

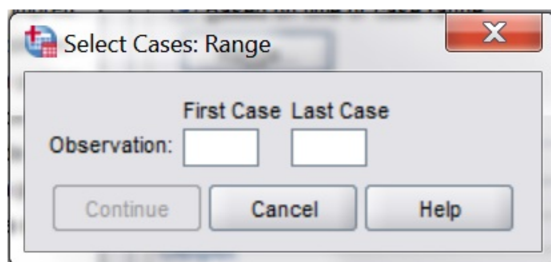
Graficul 6.69 Data, Select Cases Random sample of cases în SPSS



A doua variantă de extragere a unui eșantion aleator presupune indicarea exactă a unui număr de cazuri care vor fi selectate dintre primele cazuri (în această situație se va specifica exact numărul de cazuri care vor fi selectate precum și numărul primelor cazuri care vor fi cuprinse în procesul de selecție).

Selectarea cazurilor se poate face și în funcție de un interval indicat, care se referă la cazuri, date calendaristice sau momente de timp (graficul 6.67, „Based on time or case range"). Dacă este vorba despre cazuri, numerele care trebuie indicate („First case", „Last case") sunt cele afișate în stânga bazei de date, în fereastra principală „DataEditor", varianta „Data View" (SPSS Survival manual A step by step guide to data analysis) .

Graficul 6.70 Data, Select Cases, Based on time or case range în SPSS



Intervalele de timp și de date calendaristice sunt disponibile doar pentru seturile de date definite ca serii de timp (meniul „Date", „Define Dates").

Ultima variantă de selectare a unui set de cazuri este „Use filter variable" (graficul 6.67) care presupune utilizarea unei variabile destinată special pentru procesul de selecție. Toate valorile variabilei diferite de 0 și valide (diferite de „missing") vor fi cuprinse în grupul de cazuri supus analiz ei ulterioare (Oficial IBM SPSS Statistics 21 Brief Guide).

Opțiunea „Select cases" are mai multe variante de tratare a datelor selectate, respectiv a celor neselectate. „Filter Out Unselected Cases" presupune că datele neselectate nu sunt cuprinse în analză dar rămân în baza de date și pot fi utilizate mai târziu dacă se revocă selecția („Select All Cases"). De notat că dacă selectăm un eșantion aleator sau dacă utilizăm o expresie logică pentru a delimita un set de cazuri, programul SPSS va genera o variabilă filtru având valoarea 1 pentru cazurile selectate și 0 pentru celelalte.

O altă variantă de tratare a cazurilor selectate constă în salvarea lor într-o nouă bază de date („Copy selected cases to a new dataset"). Cazurile neslectate nu vor fi incluse în fișierul nou creat iar fișierul inițial își va păstra forma de la început (conținând toate datele).

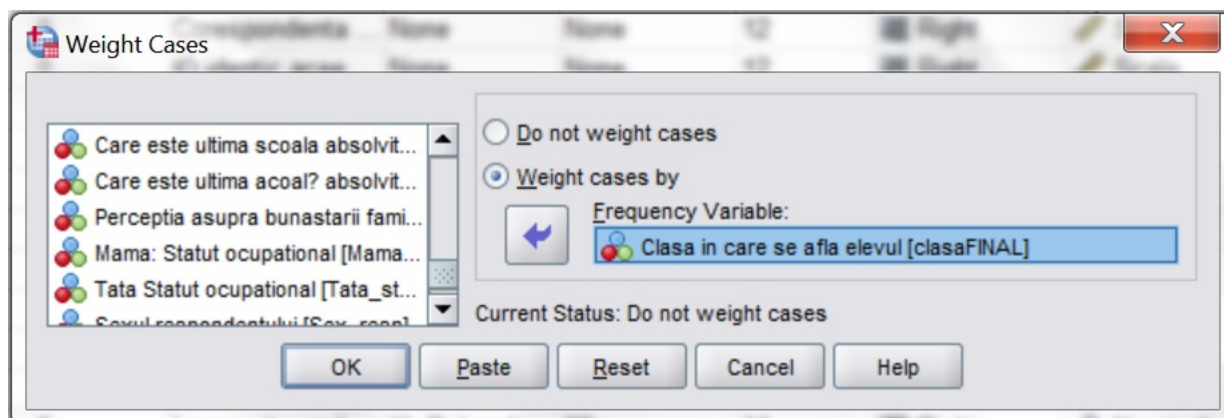
Ultima variantă de de gestionare a datelor selectate constă în ștergerea din fișierul activ a cazurilor care nu au fost selectate, fapt care face imposibilă recuperarea ulterioară a datelor neselectate. În aceste condiții, singura variantă de păstrare a datelor neselectat ar fi închiderea fișierului fără salvarea nici unei modificări și redeschiderea acestuia (Oficial IBM SPSS Statistics 21 Brief Guide).

Ultima opțiune existentă în meniul „**Data**" o constituie „**Weight Cases**" prin care fiecărui caz dintr-o bază de date i se atribuie un anumit factor de multiplicare / demultiplicare astfel încât, prin simulare, este luat în calculele statistice de un număr de ori egal cu respectivul factor. Pentru ca ponderarea să fie corectă trebuie ca valorile variabilei de ponderare să indice corect numărul de observații reprezentate de cazurile individuale din baza de date).

De exemplu, dacă am sumat două eşantioane egale de etnnici romi şi români dar ştim că, de fapt, în populația reală romii sunt de 10 ori mai puțini decât românii, cazurilor reprezentând români le atribuim ponderea 1 iar cazurilor reprezentând romi le dăm ponderea 0,1.

În graficul 6.71 e prezentată fereastra care apare când e dată comanda „Data”, „Weight Cases”. Se va selecta o variabilă de ponderare dintre cele definite în baza de date și va fi deplasată în căsuța „Frequency Variable” (Introduction to SPSS 19.0).

Graficul 6.71 Data, Weight Cases, în SPSS



E important de reținut că acele cazuri pentru cara valoarea variabilei de ponderare e 0, negativă sau dată lipsă vor fi excluse din analizele ulterioare. Valorile fracționare sunt valide, iar unele proceduri, cum ar fi calcului distribuției de frecvență sau a tabelelor de asociere, vor utiliza valori de ponderare fracționare. Totuși, majoritatea procedurilor vor trata variabilele de ponderare ca variabile de replicare a cazurilor și vor rotunji, pur și simplu, ponderile fracționare la cel mai apropiat număr întreg. Unele proceduri ignoră complet variabila de ponderare și această limitare e indicată în documentația acelei proceduri.

De notat că variabila de ponderare rămâne valabilă până când este înlocuită cu alta sau până se renunță la respectiva procedură (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

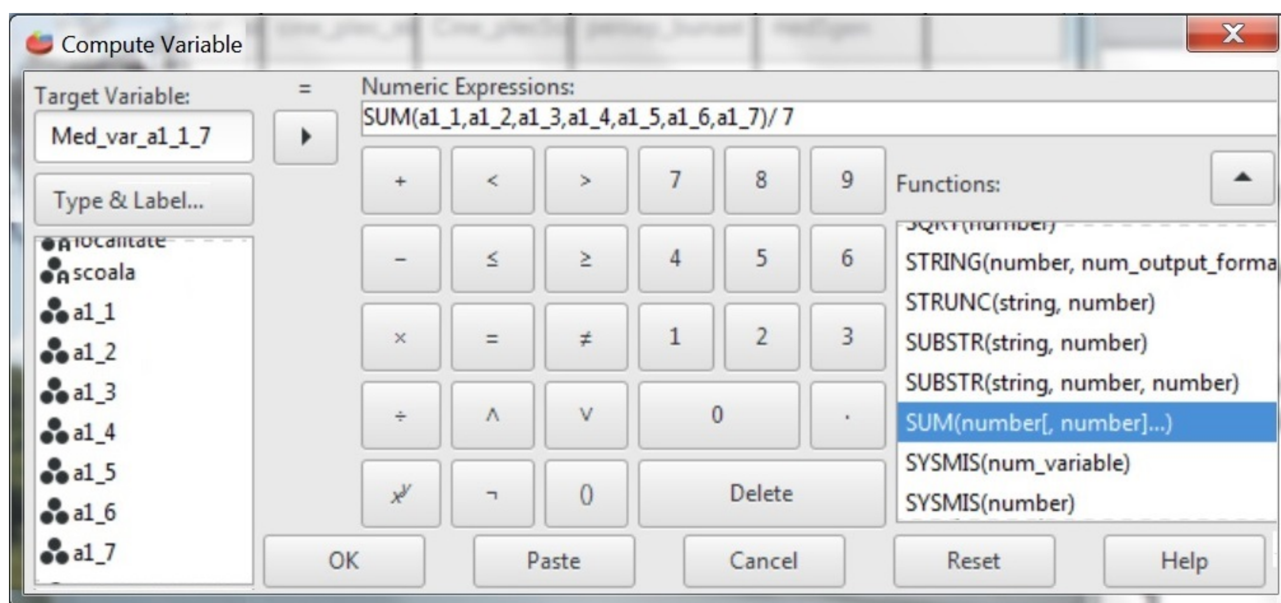
7. MENIUL TRANSFORM ÎN SPSS ȘI PSPP

7.1 Opțiunile TRANSFORM, Compute, Count Occurances of Values within Cases, în PSPP și SPSS

Începem capitolul destinat meniului „Transform” prezentând în paralel opțiunile programului PSPP și SPSS (relevând, similaritățile și diferențele).

Prin comanda „**Transform**”, „**Compute**” în programul PSPP se obțne fereastra de dialog din graficul 7.1. Numele noii variabile care va fi construită apare în stânga sus (Target variable) iar expresia aritmetică care indică modul în care se va face calculul apare în dreapta (Numeric Expressions). În partea din stâga jos a graficului 7.1 avem lista variabilelor aflate în baza de date care pot fi selectate și trimise în căsuța „Numeric Expressions”, eventual ca argumente ale unor funcții (Basic Stats in PSPP For NICAR 2016 Denver).

Graficul 7.1 Transform, Compute, în PSPP

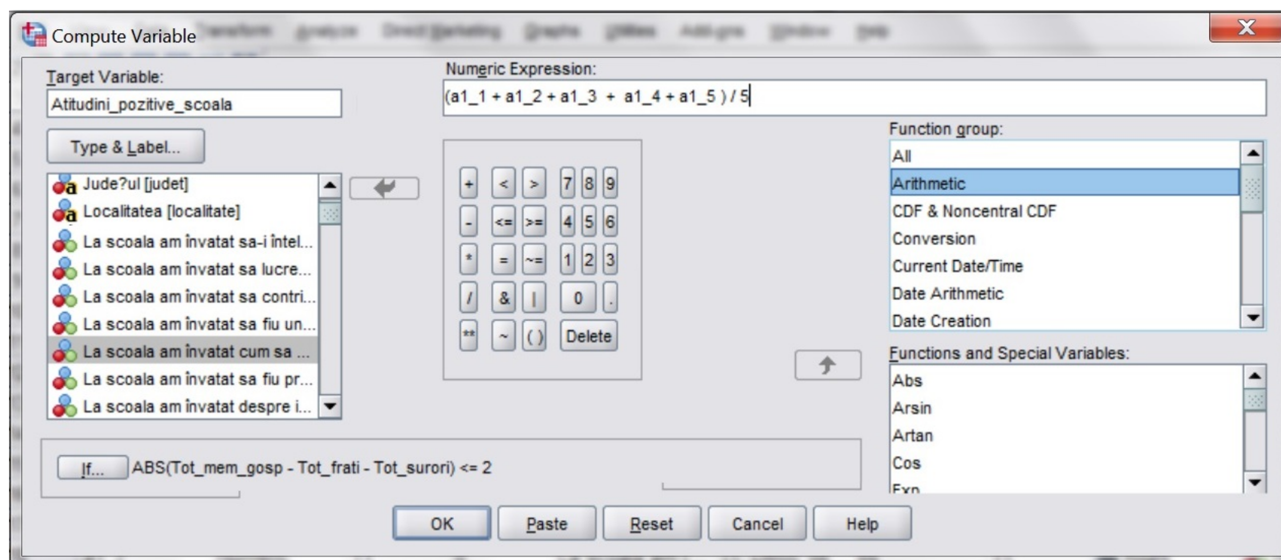


În patea din dreapta a graficului apare o listă de funcții care pot fi folosite în calculul variabilei țintă („Target variable”). În exemplul nostru am folosit variabila SUM care acceptă ca argument două sau mai multe variabile numerice (între care trebuie puse virgule). Varibilele a1_1, a1_2, ... a1_7 exprimă acordul privitor la aspecte pozitive învățate la școală (1=Deloc de acord, 2=Mai degrabă în dezacord, 3=Mai degrabă de acord și 4=Cu totul de acord) iar suma variabilelor împărțită la 7 ar exprima acordul mediu privind lucrurile bune deprinse în calitate de elev. Oricare dintre funcțiile care se dorește a fi folosite în construcția variabilei trebuie să fie selctate în preabil (din „Functions”) și trimise, dând click pe triunghiul negru, in căsuța „Numeric Expressions” (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.)

Și în programul SPSS există comanda „Transform”, „Compute” dar are unele facilități în plus. În exemplul din graficul 7.2 am folosit cinci variabile vizând lucrurile învățate la școală pe care le-am sumat aritmetic și le-am împărțit la 5. Variabila nou creată a fost denumită „Atitudini_pozitive_scoala” „Using SPSS For Windows Data Analysis and Graph”.

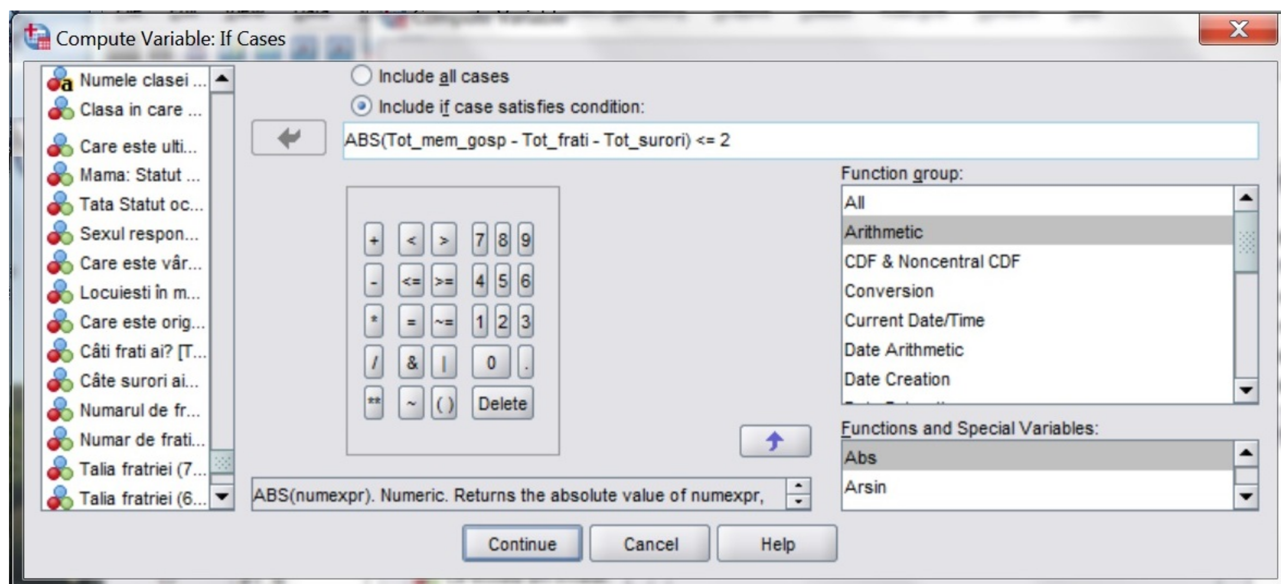
Comanda „Transform", „Compute" în SPSS diferă major de cea existentă în PSPP prin faptul că permite condiționarea calcululului variabilei țintă („Target variable") de îndeplinirea unei condiții logice.

Graficul 7.2 Transform, Compute, în SPSS



Dacă în fereastra din graficul 7.2 se dă click pe butonul „If" din stânga jos, o nouă fereastră se va deschide (graficul 7.3), având în stânga lista variabilelor din baza de date, în partea de sus expresia logică care condiționează calculul variabilei țintă și în dreapta grupuri de funcții („function groups") și conținuturile acestora („Functions and Special Variables").

Graficul 7.3 Transform, Compute, If Cases, în SPSS

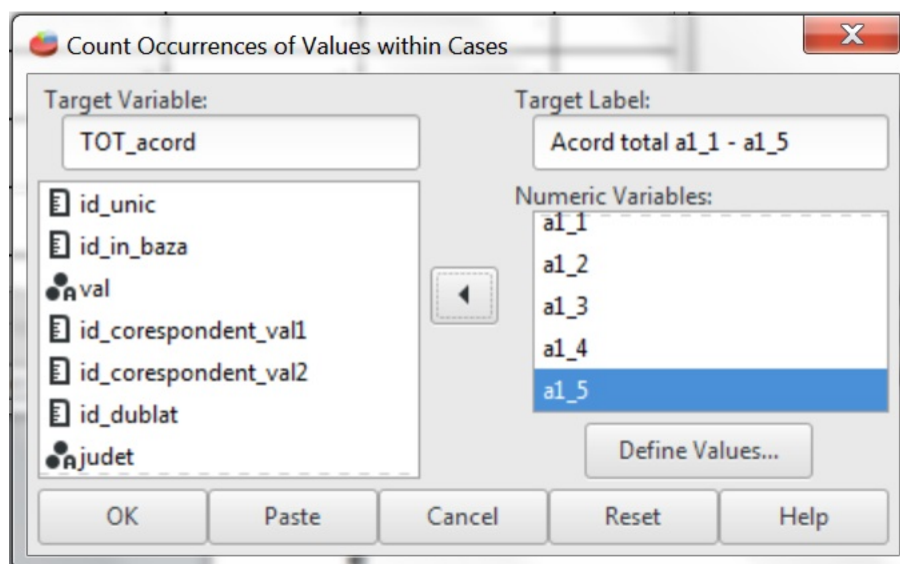


În exemplul nostru am folosit grupul de funcții „Arithmetic" și am ale din acest grup funcția ABS (returnează valoare numărului cu semn pozitiv). Am deplasat funcția ABS (folosind săgeata) în căsuța de sus și apoi am introdus expresia (Tot_mem_gosp – Tot_frati – Tot_surori) folosind lista din stânga (Using SPSS For Windows Data Analysis and Graph).

Variabila țintă „Atitudini_pozitive_scoala” va fi calculată dacă valoarea ABS (Tot_mem_gosp – Tot_frati – Tot_surori) va fi mai mică sau egală cu 2 (în general, va fi vorba de gospodăriile în care sunt cel mult 2 adulți).

Următoarea opțiune a meniului „**Transform**”, în cazul programului PSPP, este „**Count Occurances of Values within Cases**”. Facilitatea oferită este de a construi noi variabile ale căror expresii numerice rezultă din numărul de apariții ale anumitor valori pentru una sau mai multe dintre variabile (PSPP Users’ Guide GNU PSPP Release 1.0.1.).

Graficul 7.4 Transform, Count Occurances of Values within Cases în PSPP



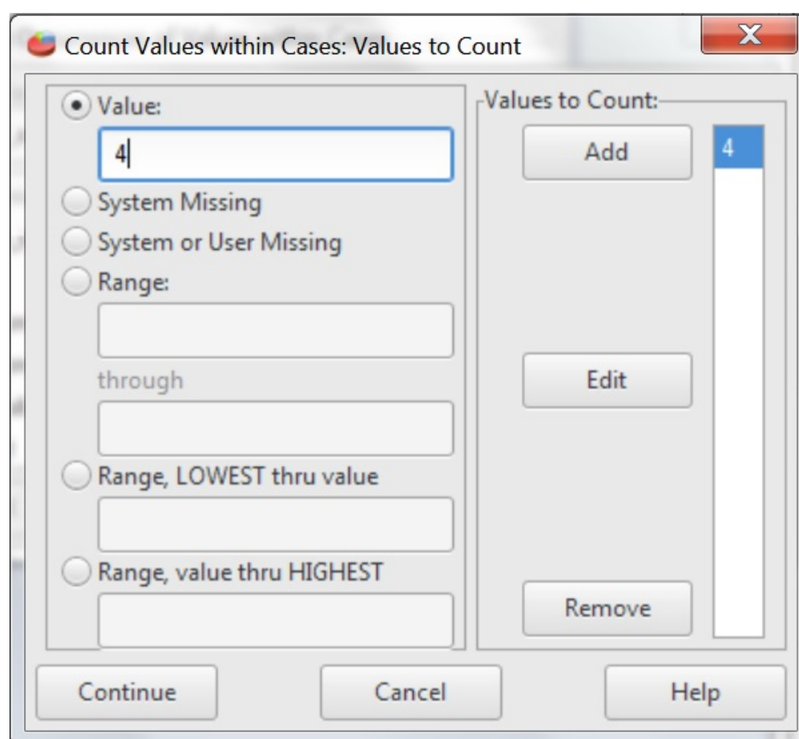
În exemplul din graficul 7.4 am definit variabila țintă „Tot_acord” (cu eticheta „Acord total a1_1 - a1_5”) desemnând numărul de răspunsuri „Total acord” la întrebările codificate a1_1, a1_2 ... a1_5 referitoare la măsura în care elevii acceptă că au deprins diferite atitudini la școală. Variabilele a1_1 - a1_5 au fost selectate din lista aflată în drepta graficului 7.4 și au fost mutate, folosind săgeata, în căsuța „Numeric variables”.

În continuare trebuie selectat butonul „Define Values” pentru a se putea specifica ce valori vor fi numărate la nivelul variabilelor selectate (PSPP Users’ Guide GNU PSPP Release 1.0.1.). În graficul 7.5 sunt prezentate modalitățile oferite de PSPP pentru construirea variabilei țintă:

- Alegând „Value” se va stabili valoarea sau valorile care să fie numărate (în exemplul nostru am optat doar pentru valoarea 4 dar puteam adăuga și 3)
- Selectând „System Misssing” se vor număra non-răspunsurile constând din lipsa completă a informației din baza de date
- Varianta „System or User Misssing” va genera o variabilă care va număra toate datele lipsă (cele definite de utilizator și „System Misssing”)
- Cu „Range” se poate indica un interval de valori care să fie numărate
- Opțiunea „Range, LOWEST thru value” va număra toate valorile începând cu cea mai mică până la cea indicată în căsuța de dedesubt
- Alegând „Range, value thru HIGHEST” se vor număra toate valorile începând cu cea indicată în căsuța de dedesubt până la cea mai mare (PSPP Users’ Guide GNU PSPP Release 1.0.1.).

De notat că în căsuța din dreapta graficului 7.5 se pot introduce mai multe cerințe, cum ar fi numărarea valorilor egale cu 4 și a datelor lipsă.

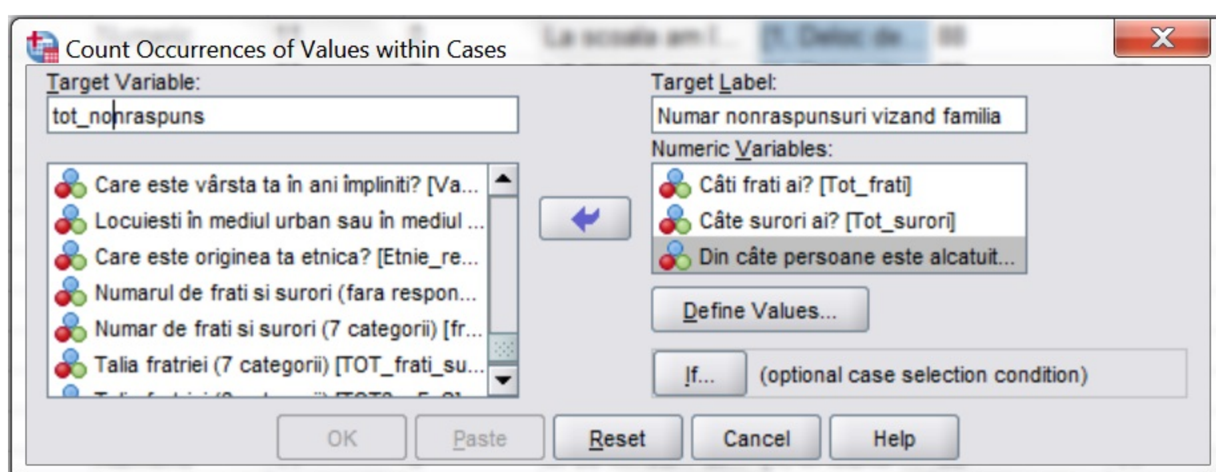
Graficul 7.5 Count Values within Cases: Values toCount în PSPP



Programul PSPP va genera o singură variabilă care va fi definită, pentru fiecare caz, prin numărul de valori ale variabilelor care corespund setului de cerințele formulate (Values of Count).

În programul SPSS modul de construcție al variabilei care numără valorile uneia sau mai multor caracteristici din baza de date este același cu cel din PSPP, comanda fiind „**Transform**", „**Count Occurence of Values within Cases**". Singura diferență notabilă este faptul că în SPSS există posibilitatea de a alege ca variabila țintă să fie calculată doar dacă e îndeplinită o condiție logică. Fereastra e identică cu cea din graficul 7.3 așa că nu reluăm explicațiile (SPSS for Beginners Copyright 1999 Vijay Gupta)

Graficul 7.6 Transform, Count Occurances of Values within Cases în SPSS

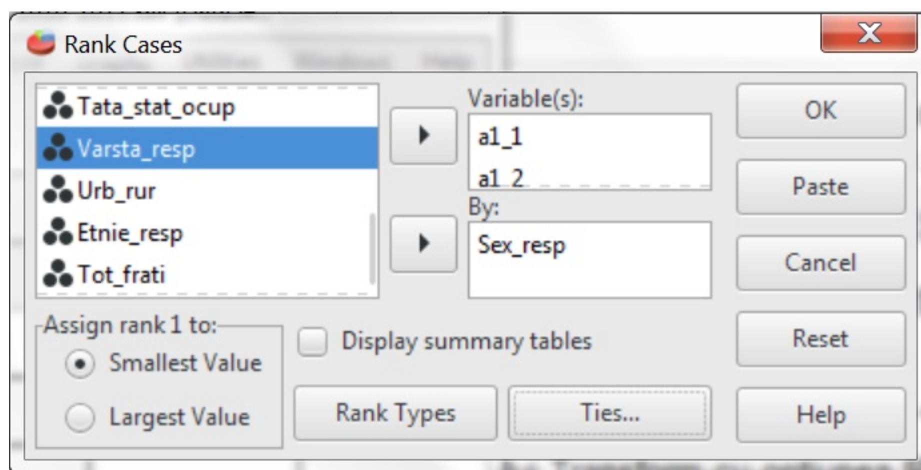


7.2 Opțiunile TRANSFORM, Rank Cases, Automatic Recode în SPSS și PSPP

Prezentăm, în continuare, opțiunea **Rank Cases**, care vizează definirea de noi variabile ale căror valori au ca punct de plecare poziția (rangul) pe care se află cazurile din variabilele inițiale (sortate ascendent sau descendent – pentru variabilele care pot fi sortate, evident).

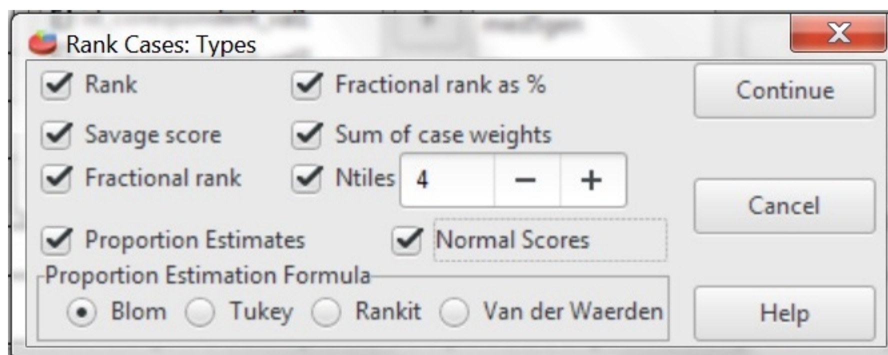
În stânga graficului 7.7 avem lista variabilelor din baza de date și dintre acestea pot fi alese una ori mai multe dar doar din cele definite ca numerice (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.)

Graficul 7.7 Transform, Rank Cases în PSPP



Se poate alege între atribuirea rangului 1 celei mai mici valori sau celei mai mari (Assign rank 1 to:). În căsuța „Variable(s)" din dreapta graficului 7.7 vor fi introduse varabilele pentru care se dorește aplicarea procedurii „Rank Cases" iar în căsuța „By:" aleg unai sau mai multe variabile categoriale care vor defini grupurile unde vor avea loc sortările. În exemplul nostru am ales două variabile (a1_1 și a1_2) pentru ale căror valori să se atribuie ranguri iar caracteristica care va defini cele două grupuri de cazuri (în care se vor fi atribuite respectivele ranguri) este variabila sex. Apoi vom alege tipul de rang pe care dorim sa-l calculăm („Rank Cases", „Types", Graficul 7.8).

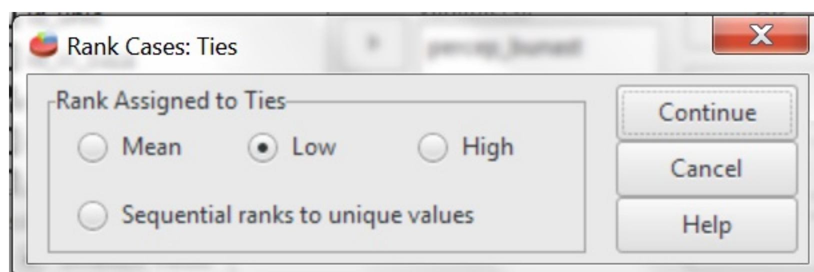
Graficul 7.8 Transform, Rank Cases, Types in PSPP



Pot fi selectate diverse metode de calculare a rangului și pentru fiecare în parte va fi generată variabila corespunzătoare (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

Opțiunea „Rank" simplă presupune că variabila nou construită va fi chiar rangul atribuit (dacă sunt mai multe cazuri cu aceeaşi valoare, rangul poate fi atribuit în moduri diferite - graficul 7.9 Rank Cases, Tyes in PSPP).

Graficul 7.9 Transform, Rank Cases, Tyes in PSPP



Dacă se alege varianta „Mean", cazurilor cu ranguri egale li se vor atribui valoari egale cu media dintre rangul anterior și cel următor. Alegând varianta „Low", cazurilor cu valori egale li se vor aloca același rang, egal cu cel inferior plus 1. Alegând varianta „High", cazurilor cu aceleași valori li se vor aloca ranguri egale, calculate scăzând 1 din rangul superior.

Varianta „Sequential ranks to unique values" presupune faptul că rangurile se atribuie în ordine cazurilor din grup iar pentru cele cu aceleași valori, rangurile sunt identice. Pentru ranguri se folosesc în ordine toate numerele naturale, spre deosebire de „Mean", „Low" sau „High" (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

Tabelul 7.1 Alocarea rangurilor în PSPP

Valoare	Mean	Low	High	Sequential
3	1	1	1	1
5	2	2	2	2
10	4	3	5	3
10	4	3	5	3
10	4	3	5	3
20	6	6	6	4

Revenind (graficul 7.8), există următoarele opțiuni de a defini a rangurile:

- „Savage score", situație în care noua variabilă conține scoruri bazate pe o distribuție exponențială.
- „Fractional rank", pentru care valoarea noii variabile este egală cu rangul împărțit la suma ponderilor cazurilor valide
- „Fractional rank as percent" presupune că fiecare rang este împărțit la numărul de cazuri valide și apoi e înmulțit cu 100
- „Sum of case weights" e o procedură prin care noii variabile i se atribuie suma ponderilor cazurilor din interiorul unui grup (valoarea va fi aceeași pentru toate cazurile din acel grup)
- „Ntiles" presupune clasamente bazate pe grupuri percentile, fiecare având aproximativ același număr de cazuri (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

De exemplu, valoarea 10 pentru „Ntiles” ar atribui rangul 1 primelor 10% valori, rangul 2 următoarelor 10% valori,etc., și rangul 10 ultimelor 10%.

Selectând „Proportion estimates”, se vor estima valorile care corespund distribuției cumulative proprii unui anumit rang.

Optând pentru „Normal Scores” vom calcula scorul „Z” care corespunde distribuției cumulative estimate. Dacă distribuția cumulativă estimată este egală cu 0.5, scorul va fi egal cu 0.

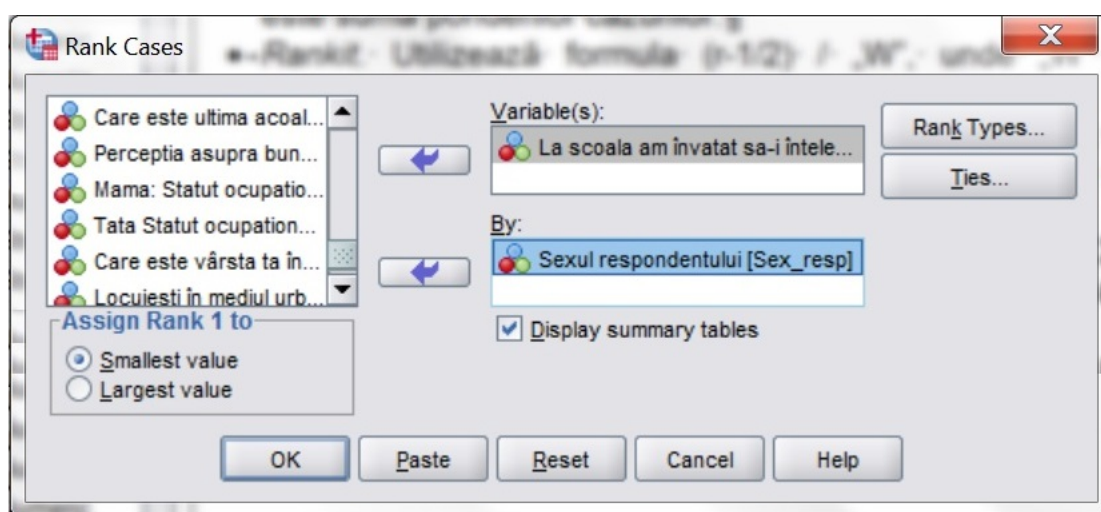
„Proportion Estimation Formula” permite ca pentru estimarea proporțiilor și scorurilor normale, să poată fi folosită una din formulele: Blom, Tukey, Rankit sau Van der Waerden.

Semnificația formulelor e următoarea:

- *Blom*. Creează o variabilă de clasificare bazată pe estimările proporției care este definită prin formula (r-3/8) / (w + 1/4), unde „W” e suma ponderilor cazurilor și „R” este rangul.
- *Tukey*. Utilizează formula (r-1/3) / (w + 1/3), unde „R” este rangul și „W” este suma ponderilor cazurilor.
- *Rankit*. Utilizează formula (r-1/2) / „W”, unde „W” este numărul de observații și „R” este rangul, variind de la 1 la „W”.
- *Van der Waerden*. Transformarea lui Van der Waerden, este definită de formula r / (w + 1), unde „W” este suma ponderilor cazurilor și „R” este rangul, variind de la 1 până la „W” (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1.).

În programul SPSS opțiunea **Transform**, **Rank Cases** este practic identică cu cea din PSPP astfel încât respectiva opțiune o vom descri-o în doar câteva cuvinte (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

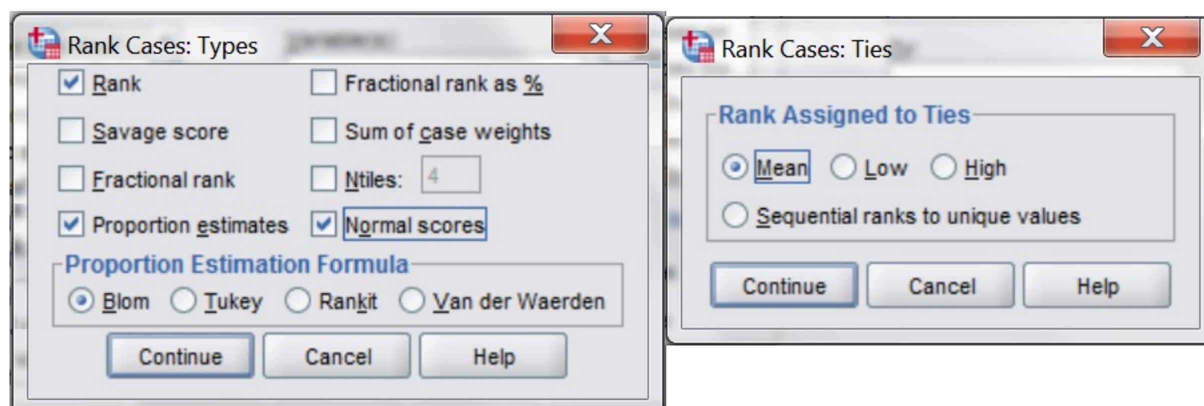
Graficul 7.10 Transform, Rank Cases în SPSS



Graficul are aceeași structură ca și cel din programul PSPP (graficul 7.7) cu deosebirea minoră că butonul „Rank Types” (care permite selectarea modului de calcul al rangului) și butonul „Tyes” (care stabilește regula de succesiune a valorilor rangului) se află în colțul din dreapta sus și nu în partea de jos a ferestrei.

Vom prezenta ferestrele analoage celor care apar în PSPP („Rank Types” și „Tyes”) dar fără să mai detaliem procedurile programului SPSS (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

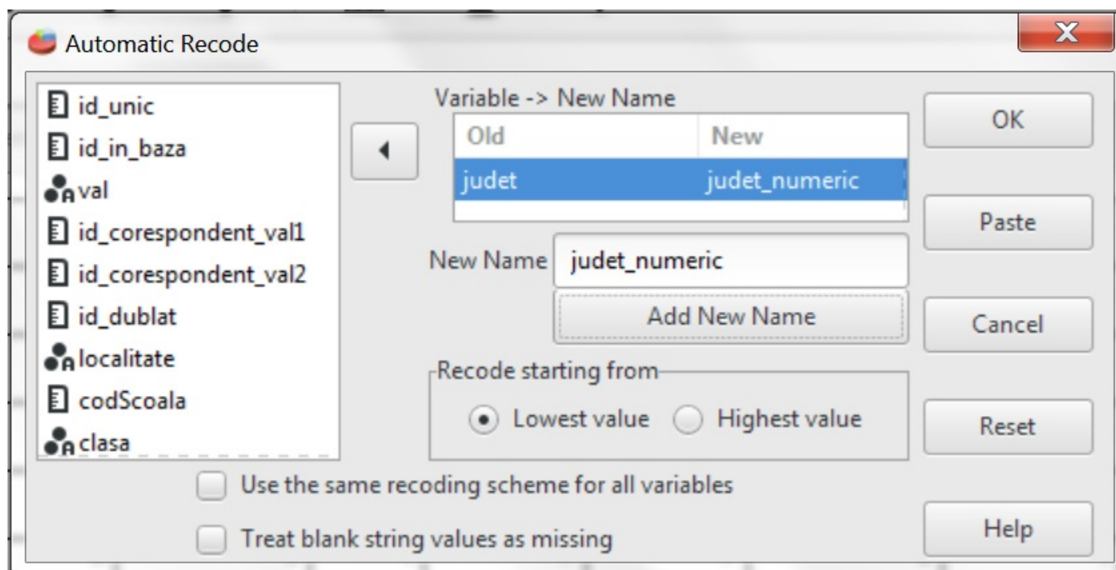
Graficul 7.11 Transform, Rank Cases, ferstrele Types și Ties în SPSS



Având în vedere că gaficul 7.11 conține exact aceleași butoane și comenzi ca și graficele 7.8 și 7.9 (care prezintă „Rank Cases", „Types" respectiv „Ties" în PSPP), vom adăuga doar că trebuie urmate exact aceleași proceduri ca și în PSPP.

Următoarea opțune a meniului **Transform**, în PSPP, este „**Automatic Recode**" și presupune atribuirea de valori întregi succesive (începând cu 1) grupurilor de valori identice ale variabilelor (string sau numerice) în funcție de poziția acelor grupuri în ordinea lor crescătoare sau descrescătoare (PSPP Users' Guide GNU PSPP Release 1.0.1.)

Graficul 7.12 Transform, Automatic Recode în PSPP

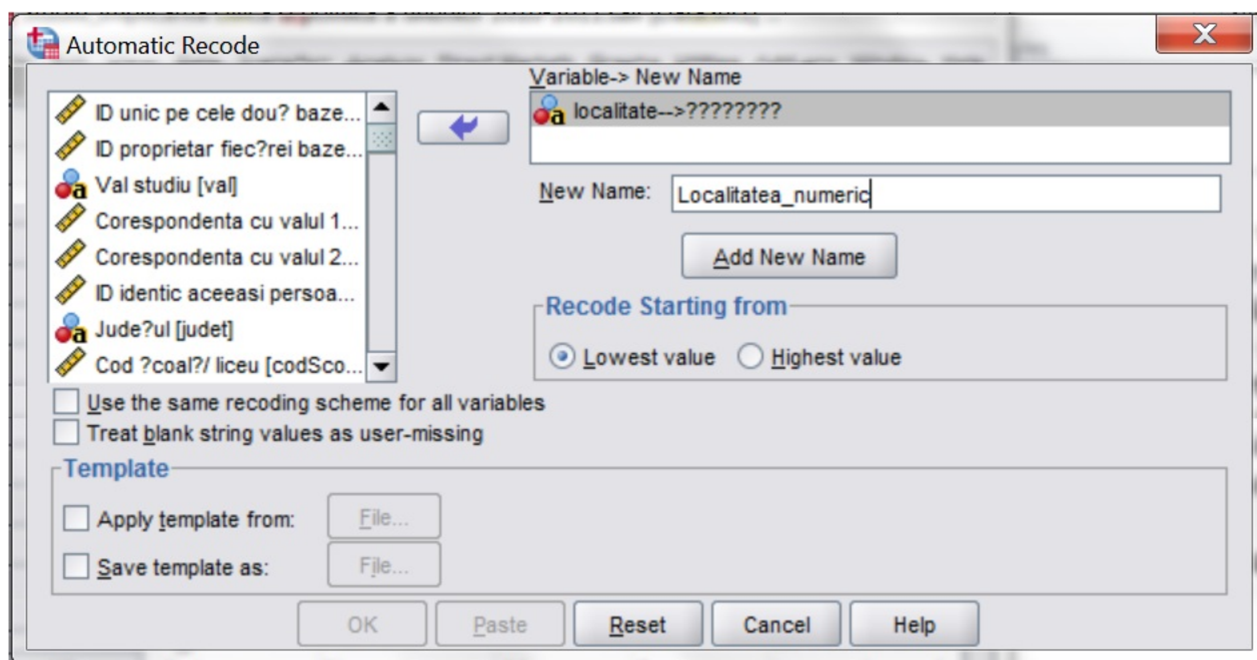


Procedura este utilă mai ales pentru faptul că unele analize din PSPP nu acceptă variabile „string" iar altele au nevoie de valori întregi pentru estimarea nivelului unor anumiți factori. Datele lipsă nu sunt recodate ca „Missing" dar iau valori mai mari decât oricare dintre valorile valide recodificate. Variabila nou creată va avea ca etichete ale valorilor numerele sau șirurile de caractere care au fost proprii fiecărui caz iar eticheta numelui variabilei nu va fi preluată de la cea recodificată. E important faptul că recodificarea se poate face în ordine crescătoare sau descrescătoare folosind căsuța de dialog „Recode starting from" (PSPP Users' Guide GNU PSPP Release 1.0.1.)

Opțiunea „**Automatic Recode**” dă posibilitatea recodificării mai multor variabile conform aceleiași proceduri („Use the same recoding scheme for all variables”). Limitările impuse sunt ca toate variabilele să fie de același tip (numeric sau string). De notat că datele lipsă definite de utilizator (User-missing) vor fi tratate ca date valide în noile variabile create. Menționăm faptul că, în mod normal, pentru variabilele „string” datele care efectiv lipsesc nu sunt tratate ca „system-missing”. Opțiunea „Treat blank string values as user-missing” va duce la autorecodificarea datelor lipsă „string” drept cea mai mare valoare dintre cele obținute prin recodificare (PSPP Users’ Guide GNU PSPP Release 1.0.1.).

Vom prezenta acum opțiunea „Automatic Recode” în varianta programului SPSS care prezintă doar câteva deosebiri față de PSPP.

Graficul 7.13 Transform automatic recode în SPSS



Vom încerca să relevăm similitudinile și deosebirile care apar în SPSS față de programul PSPP. În primul rând noile variabile create prin „Automatic Recode” preiau etichetele variabilelor și valorilor acestora și le aplică valorilor nou calculate. Pentru orice nume de variabilă sau valori care nu au etichete, valorile originale vor fi folosite drept etichete ale valorilor recodificate (în fișierul output vor fi listate vechile etichete și cele noi, care le corespund). De reținut că variabilele string sunt recodificate în ordine alfabetică atât în programul SPSS cât și în PSPP (Introduction to SPSS 19.0).

Pentru variabilele numerice datele lipsă sunt recodificate la valori mai mari decât cea mai mare valoare recodificată validă.

Pentru variabilele de tip „string” valorile lipsă sunt recodificate la valoarea 1 dacă nu e selectată opțiunea „Treat blank string values as user-missing”. Dacă opțiunea e selectată, datele lipsă sunt recodificate la o valoare superioară celei mai mari valori valide a variabilei nou-creată.

Facilitatea „Use the same recoding scheme for all variables” e identică cu cea din PSPP, premițând recodificarea automata a mi multor variabile simultan (IBM SPSS Statistics 20 Core System, User’s Guide).

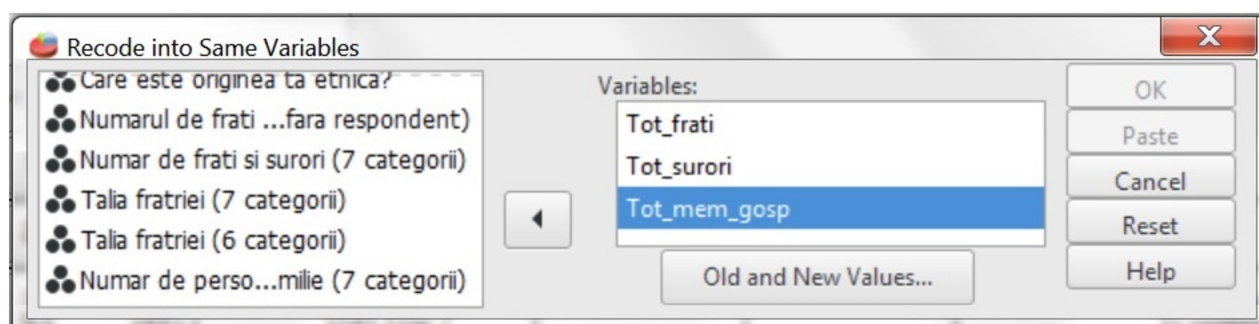
Reiterăm ideea că variabilele trebuie să fie de același tip (numerice sau „string” iar valorile considerate „user missing” vor fi cele întâlnite la prima dintre variabilele din grupul celor supuse reodificării.

Programul SPSS are o opțiune în plus față de PSPP pentru „Automatic recode”, intitulată „Templates” prin care se poate salva o schemă de recodificare automată într-un fișier „șablon” pentru a fi aplicată ulterior altor variabile / baze de date. De exemplu se poate crea un „Template” pentru datele de sondaj pre-electorale care culeg informații lunar din aceleași localități. Totuși, s-ar putra ca în anumite sondaje să fie adăugate noi localități, ceea ce ar determina modificarea valorilor obținute prin recodificare. Dacă se aplică „Template-ul” asupra datelor conținând noi localități, vechile valori de recodificare se vor păstra iar pentru noile localități se vor aloca numere superioare celor mai mari valori valide obținute prin aplicarea respectivului „Template”. Fișierul „Template”, cu extensia „.sat” poate fi salvat folosind „Save template as” și poate fi apelat (după ce s-a indicat variabila de recodificare și noul ei nume), utilizând instrucțiunea „Apply template from” (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

7.3 Opțiunile TRANSFORM, Recode into Same Variable, Recode into Different Variable în SPSS și PSPP

Revenind la programul PSPP, urmăroarea opțiune este **Transform Recode into Same Variable** care oferă posibilitatea operării unor modificări asupra valorilor, de exemplu recodificarea unei variabile continue într-una categorială prin transformarea intervalelor în valori numerice sau reducerea numărului de categorii ale unei variabile prin comasarea celor apropiate ca semnificație (Basic Stats in PSPP For NICAR 2016).

Graficul 7.14 Transform, Recode into Same Variable în PSPP

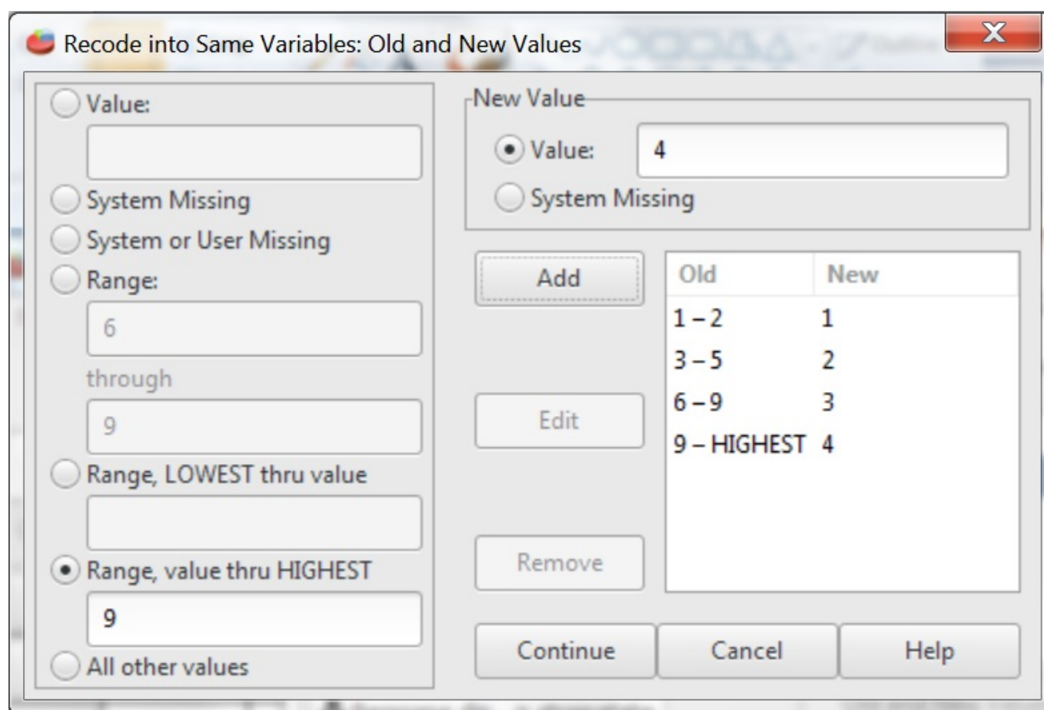


Asupra variabilelor „string” se poate aplica „Recode into Same Variable” dar numai în sensul atribuirii unor noi valori variabilei recodificate (celelalte opțiuni sunt accesibile doar variabilor numerice). De notat că putem utiliza, simultan, mai multe variabile (sau numerice, sau „string”) dar numai în cazul în care modificările pe care dorim să le realizăm prin recodificare sunt de aceeași natură pentru fiecare variabilă în parte.

Pentru recodificarea concretă a uneia sau mai multor variabile trebuie mai întâi ca acestea să fie selectate și trecute în căsuța „Variables”. În exemplul nostru am ales trei variabile numerice desemnând numărul de frați ai subiectului, numărul de surori și totalul membrilor în gospodărie (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

Următorul pas constă în comanda „Old and new Values" care deschide fereastra de dialog „Recode into same variables: Old and new values" (Basic Stats in PSPP For NICAR 2016 Denver).

Graficul 7.15 Recode into same variable Old and new values in PSPP



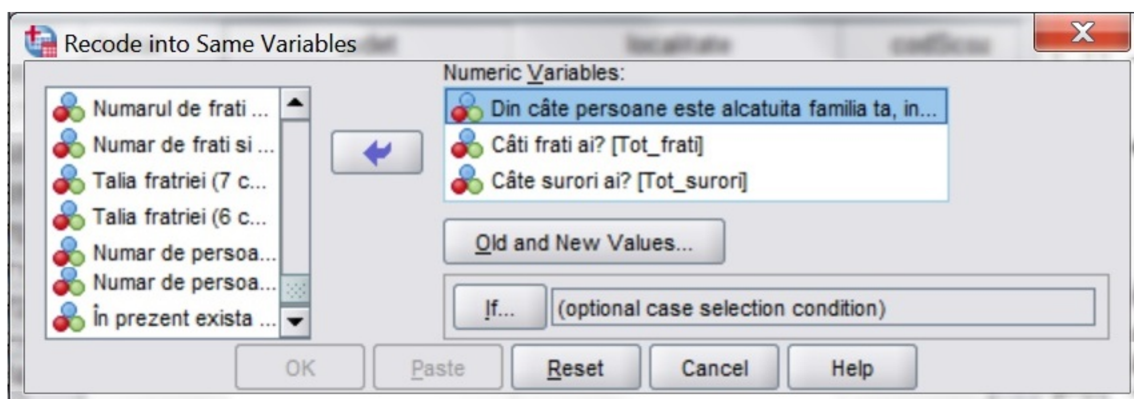
Pentru variabilele cantitative există mai multe opțiuni de operare cu valorile variabilei ce urmează a fi recodificată:

- „Value" permite selectarea unei valori a vechii variabile pentru care vom indica, apoi, noua valoare corespunzătoare în căsuța „New Value" (pe aceeași cale putem defini ca o valoare validă a vechii variabile să fie eliminată din cea recodificată, devenind „System Missing")
- „System Missing" și „System or User Missing" permit alocarea de valori numerice, în cazul varibilei recodificate, pentru datele care au fost definite ca fiind lipsă de către utilizator respectiv datelor lipsă din baza de date.
- „Range" dă posibilitatea alegerii unui interval de valori ale vechii variabile pentru care (în căsuța „New value") urmează să se atribuie o anumită valoare
- „Range LOWEST thru value" facilitează înlocuirea valorilor vechii variabile care sunt mai mici sau egale decât o valoare definită – cea introdusă în căsuța – cu o nouă valoare indicată în căsuța „New value"
- „Range value thru HIGHEST" permite înlocuirea valorilor vechii variabile care sunt mai mari sau egale cu o anumită valoare (cea introdusă în căsuță) – cu o nouă valoare indicată în căsuța „New value"

Notăm faptul că fiecare transformare pe care dorim s-o realizăm în interiorul valorilor variabilei va necesita comanda „Add" și că în căsuța din dreapta ferestrei de dialog vor fi listate toate respectivele transformări (va apărea un tabel cu vechile și noile valori ale variabilei). În exemplul nostru, intrvalelor 1-2, 3-5 și 6-9 (persoane) li se atribuie valorile 1, 2 și 3. Pentru situația în care sunt 9 sau mai multe persoane, se atribuie valoarea 4 (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

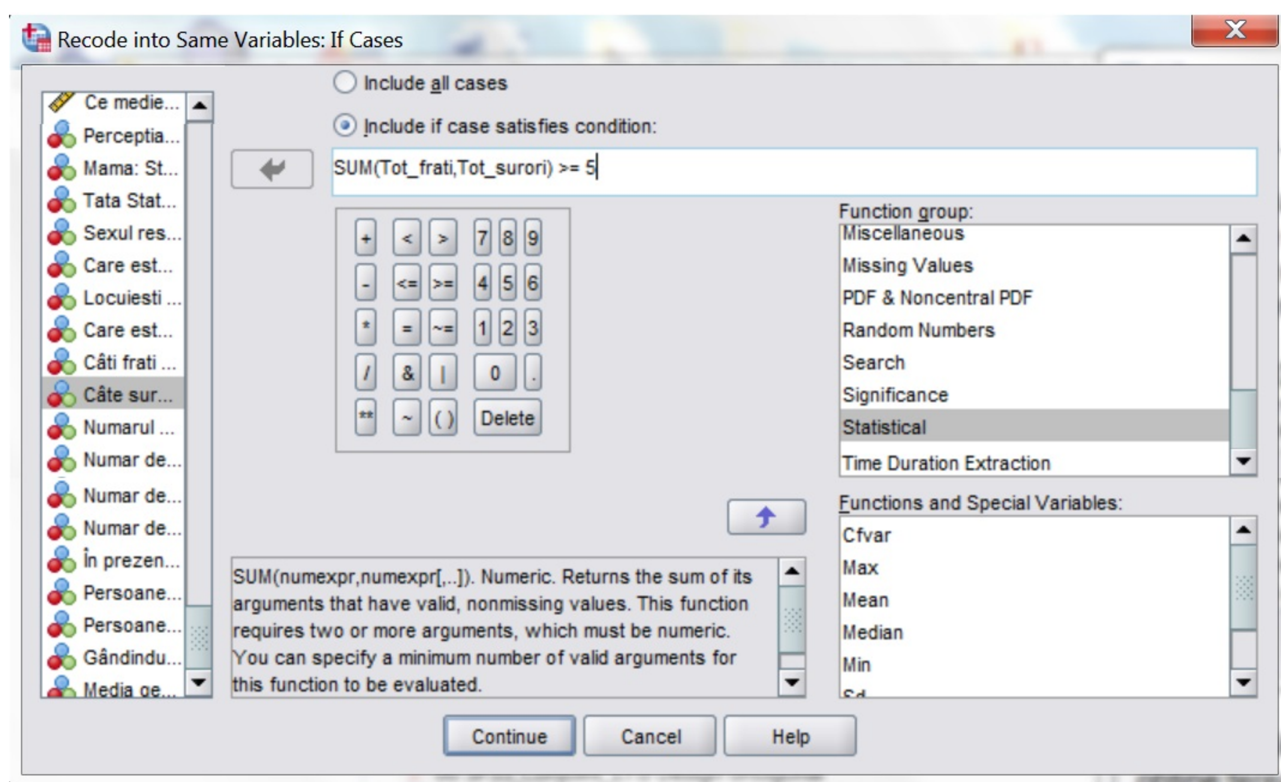
Instrucțiunea „Recode into Same Variables" din SPSS seamănă cu cea din PSPP astfel încât vom prezenta doar ferestrele de dialog aferente precum și unica diferență notabilă: recodificarea valorilor variabilelor poate fi condiționată de îndeplinirea unei condiții logice (butonul „If").

Graficul 7.16 Transform recode into same variable in SPSS



Dacă alegem să folosim condiția logică și selectăm butonul „If", vom obține fereastra de dialog din graficul 7.17.

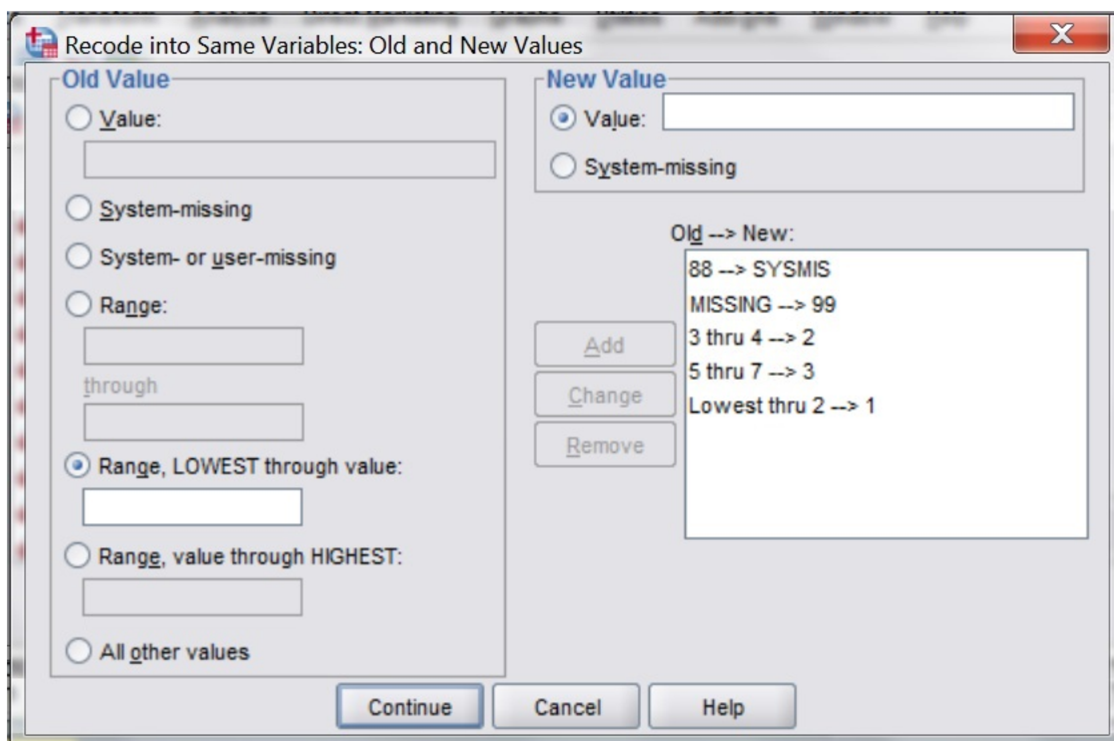
Graficul 7.17 Transform Recode into Same Variables, If Cases în SPSS



Fereastra e identică cu cea din graficul 7.3 și permite folosirea unei funcții matematice (SUM, în exemplu) și a variabilelor pentru a defini o condiție logică. Recodificarea se va face doar dacă sunt cel puțin 5 frați și surori în familie (Introduction to SPSS - version 23 - for Windows, Practical workbook).

Dacă în graficul 7.16 vom da click pe butonul „Old and new values” vom deschide fereastra prezentată în graficul 7.18.

Graficul 7.18 Recode into Same Variables, Old and new values în SPSS



În exemplul din graficul 7.18 valorile 88 ale celor 3 variabile supuse recodificării au fost transformate în date lipsă din baza de date (System missing). Pe de altă parte, datele lipsă existente la nivelul celor trei variabile (definite de utilizator sau lipsind efectiv din baza de date) sunt recodificate în valoarea 99. Valorile variabilelor cuprinse între trei și patru vor fi transformate în valoare doi iar cele între 5 și 7, în 3. Valorile între cea mai mică și 2 vor deveni 1 (Introduction to SPSS 19.0).

De notat că atunci când se recodifică mai multe variabile simultan este necesar ca acestea să aibă aceeași natură (toare numerice sau toate „string”). În cazul variabilelor „string” singura opțiune activă în fereastra 7.18 este cea de schimbare a conținutiului unei valori („Old Value, Value” – „New Value, Value”) și de atribuire a unei valori numerice pentru datele „System or User Missing” din vechile variabile.

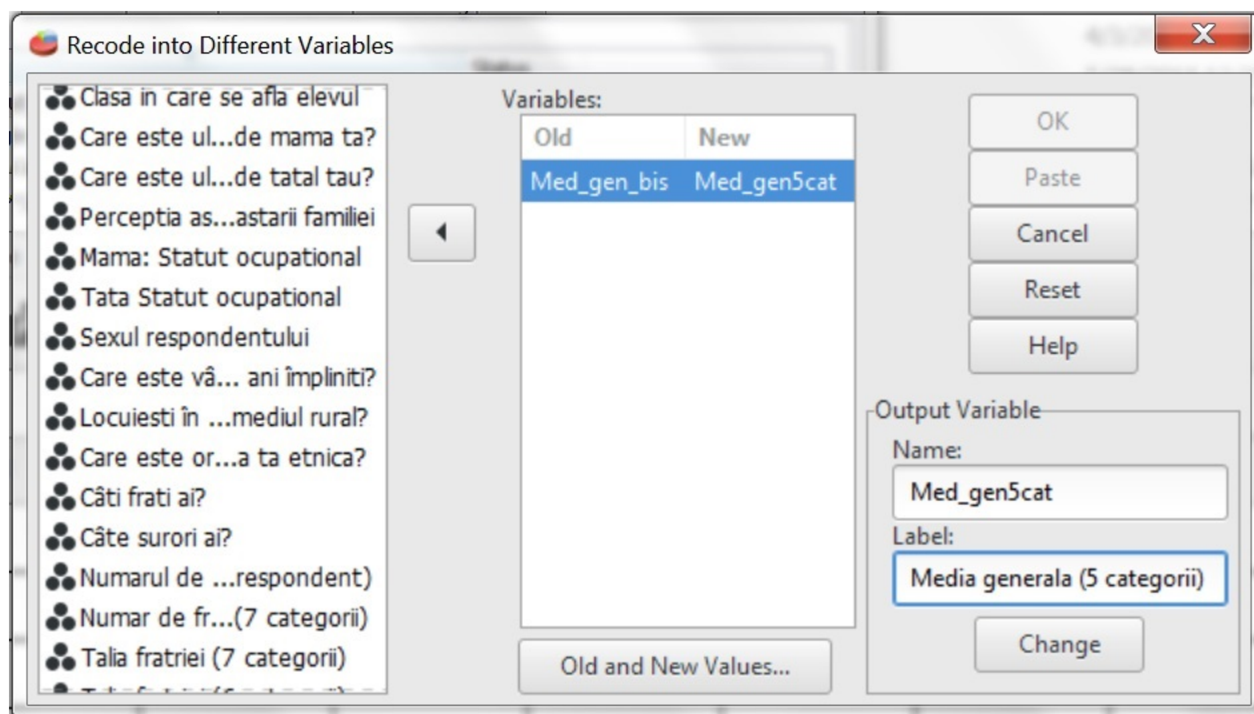
Cea mai importantă observație legată de comanda „Transform Recode into Same Variables”, indiferent că e vorba de programul PSPP sau SPSS este faptul că vechile variabile nu se mai păstrează în baza de date deci pot fi pierdute informații prețioase prin această procedură.

O opțiune asemămătoare dar care păstrează intacte variabilele utilizate în operația de recodificare este **Transform Recode into Different Variables**.

În varianta oferită de programl PSPP procedura se realizeză prin fereastra de dialog prezentată în graficul 7.19, unde în dreapta se află lista vechilor variabilele din baza de date iar la mijloc apar variabilele care vor fi recodificate (Variables, Old) și cele care le vor corespunde - Variables, New (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

În exemplul nostru am ales variabila Med_gen_bis, reprezentând media generală a anului trecut, pe care am dorit s-o recodificăm într-o nouă variabilă, „Med_gen5cat” având eticheta „Media generală (5 categorii). Numele și eticheta noii variabile trebue trecute în căsuța „Output Variable” („Name” și „Label”) după care trebuie dată comanda „Change” (pentru ca numele noii variabile sa apară în dreptul celei vechi).

Graficul 7.19 Transform Recode into Different Variable în PSPP



Pentru ca recodificarea să aibă loc este nevoie, în plus, să indicăm modul în care trebuie realizată, dând click pe butonul „Old and new Values” și lansând fereastra de dialog prezentată în graficul 7.20 (Basic Stats in PSPP For NICAR 2016 Denver).

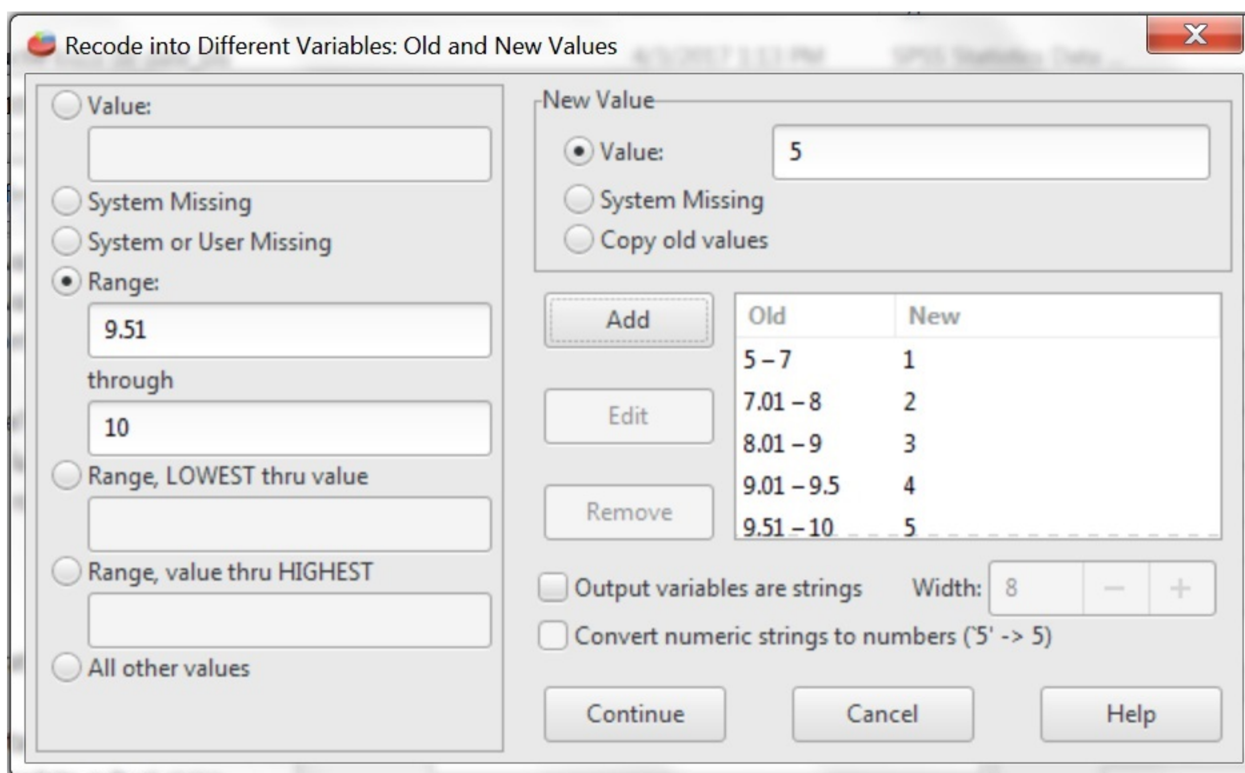
Menționăm, de la început, faptul că s-ar fi putut recodifica, simultan, mai multe variabile, dacă acestea aveau aceeași natură (ambele numerice sau ambele string) și dacă, în plus, natura operației de recodificare era aceeași. În cazul nostru, am fi putut recodifica mediile generale ale elevilor pentru clasele 9-12, utilizând aceeași procedură:

- Pentru notele cuprinse înre 5 și 7 am alocat noii variabile valoarea 1
- Pentru notele din intervalul înre 7.01 și 8 valoarea corespunzătoare e 2
- Mediilor între 8.01 și 9 li s-a atribuit valoarea 3
- Intrevalul 9.01-9.50 a fost recodificat în valoarea 4
- În final mediilor între 9.51 și 10 le-a fost atribuită valoare 5

Dacă variabila „Med_gen_bis” ar fi avut date lipsă declarate, acelea ar fi trebuit să apară și în noua variabilă, tratate în același mod (de exemplu, pentru „Value” = 99 desemnând date declarate lipsă de utilizator pentru vechea variabilă am putea să atribuim valoarea 88 („New Value”, „Value”). Ulterior va trebui, evident, să definim etichetele și datele lipsă ale noii variabile De exemplu, valorii 1 putem să-i atribuim eticheta „5-7”, valorii 2, eticheta „7.01-8”, etc. (PSPP Users’ Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

Nu în ultimul rând, opțiunea „All other values" permite ca tuturor valorilor care nu au fost cuprinse într-o procedură de recodificate să li se poată atribui o aceeași valoare (în căsuța „New Value") sau să fie copiate sub aceeași formă ca în vechea variabilă („New Value", „Copy old values"). De fapt opțiunea „Copy old values" din „New Value" poate fi aplicată orcăror valori sau intervale de valori care au fost selectate în stânga graficului 7.20.

Graficul 7.20 Recode into Different Variable Old and New Values in PSPP



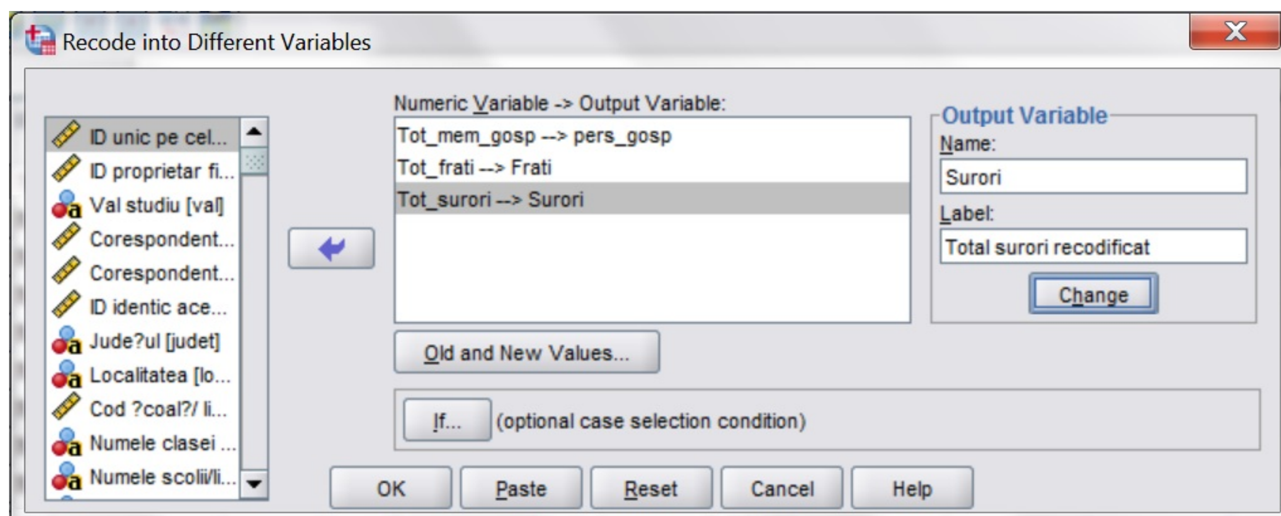
De notat că în cazul variabilelor „string", singura opțiune disponibilă din partea stângă a graficului 7.20 este „Value" iar în drepta, „New Value" (se pot, deci, schimba valorile variabilei string cu unele noi). În plus, alegând opțunea „Output variables are strings" se poate stabili numărul maxim de caractere. Mai menționăm faptul că introducând în „Value" o valoare a variabilei string, ea poate fi copiată în noua variabilă prin opțiunea „Copy old values" (PSPP Users' Guide GNU PSPP Statistical Analysis Software Release 1.0.1).

Vom prezenta în cele ce urmează procedura **Transform Recode into Different Variable** in SPSS, cu observația că există o singură diferență înportantă față de PSPP, faptul că operațiile de recodificare pot fi condiționate de îndeplinirea unei cerințe logice.

În exmplul din graficul 7.21 (unde e prezentată fereasta „Transform" „Recode into Different Variable" in SPSS) am ales un exemplu de recodificare a trei variabile. Respectivele caracteristici au fost inițial selectate din lista existentă în stânga graficului 7.21, după care au fost atribuite nume pentru noile variabile precum și pentru etichete aferente (căsuța „Outpt Variable). În final s-a dat comanda „Change".

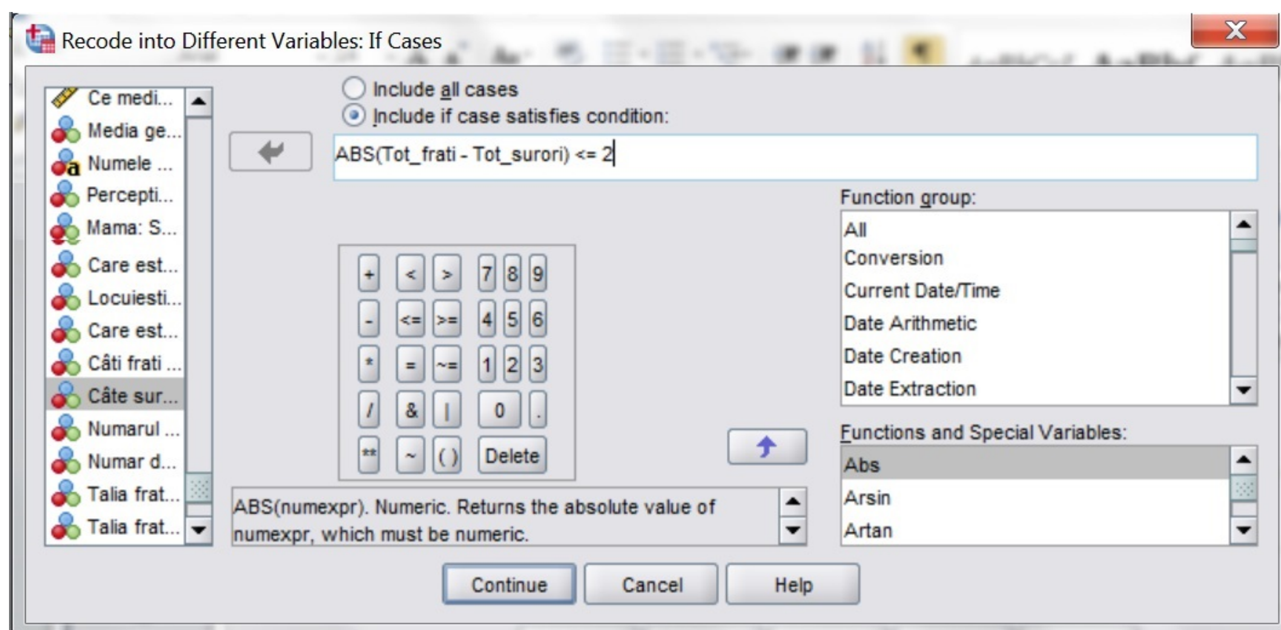
Dacă dorim ca recodificarea variabilelor să aibă loc numai în anumite condiții, putem să apelăm la opțiunea „if" („optional case selection condition") care este practic identică cu ferestrele prezentate în graficele 7.3 și 7.17. (SPSS for Windows Version 19.0 A Basic Tutorial)

Graficul 7.21 Transform Recode into Different Variable în SPSS



În graficul 7.22 este ilustrată ferastra ce permite formularea expresiei logice care condiționează recodificarea. În exemplul nostru, am pus condiția ca valoarea absoltă a diferenței dintre numărul de frați și surori să fi mai mică sau egală cu 2 (am exclus de la recodificare, astfel, familiile în care numărul fraților depășea cu mai mlt de 2 pe cel al surorilor sau viceversa).

Graficul 7.22 Transform, Recode into Different Variable, If Cases în SPSS

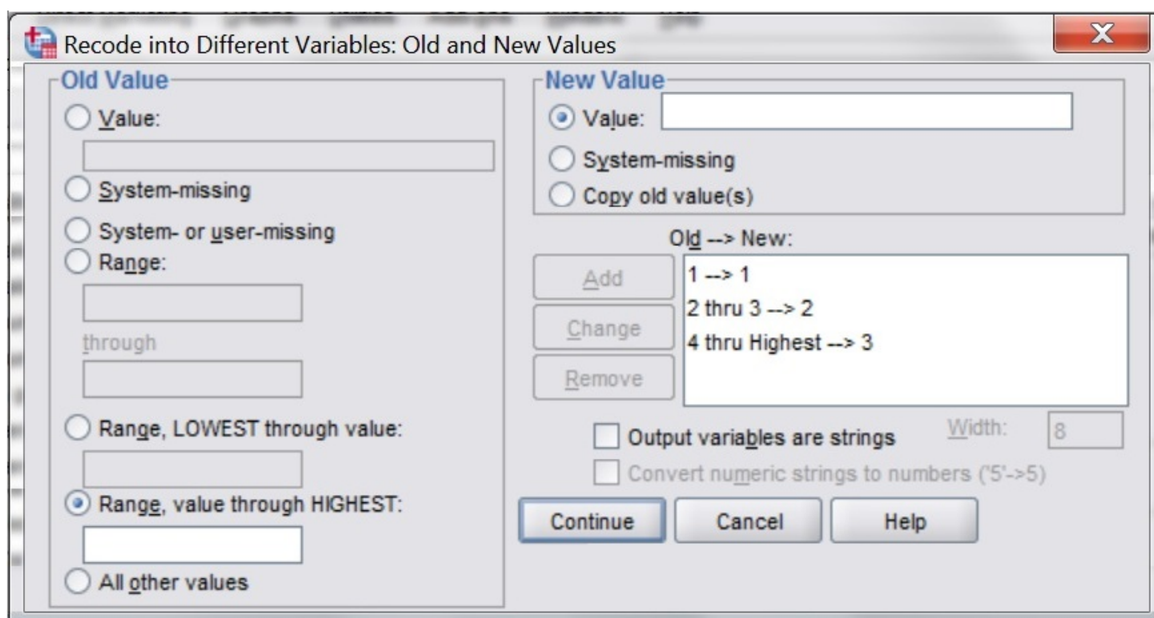


Reamintim că în fereastra din graficul 7.22, care permite formularea condiției logice, în stânga sunt listate variabilele din baza de date, la mijloc sunt plasate butoanele cu operatorii aritmetici și logici iar în dreapta apar grupurile de funcții și funcțiile aferente care pot fi folosite la construcția expresiei logice.

În exemplul nostru a fost ales (din Function group) grupul de funcții aritmetice și dintre acestea a fost selectată valoarea absolută („ABS”) din căsuța „Function and special Variables” (valoarea absolută, descrisă în partea de jos a graficului). După formularea condiție logice se va da comanda „Continue” (Using SPSS For Windows Data Analysis and Grap).

Prezentăm, în continuare, fereastra obținută prin comanda „Old and new values” inițiată în fereastra de dialog ilustrată în graficul 7.23.

Graficul 7.23 Recode into Different Variable Old and New Values în PSPP

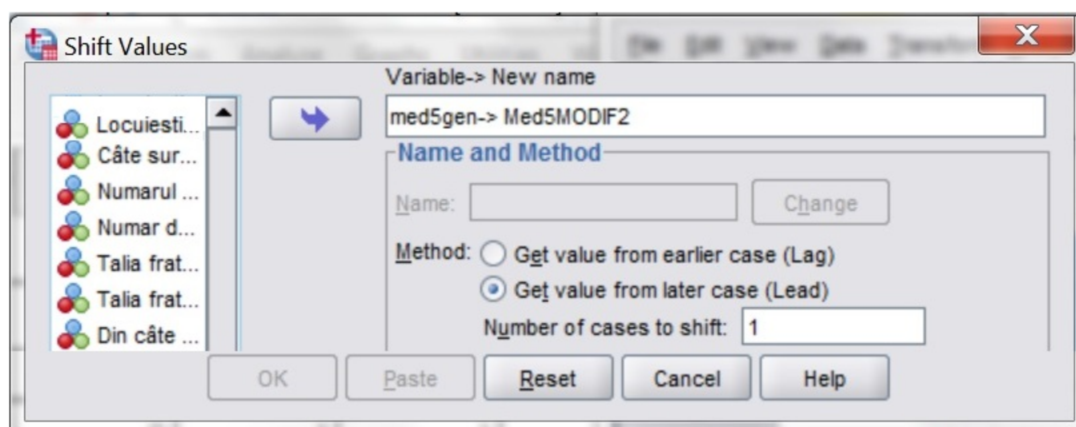


Recodificarea constă în păstrarea valorii 1 pentru variabilele în discuție (numărul de frați, surori și totalul membrilor familiei), alocarea valorii 2 pentru valorile 2 sau 3 ale variabilelor inițiale și transformarea vechilor valori mai mari sau egale cu 4 în noi valoari egale cu 3. Opțiuniile sunt identice cu cele din programul PSPP așa că nu mai continuăm explicațiile. A se vedea graficul 7.20 (Using SPSS For Windows Data Analysis and Grap).

7.4 Opțiunile TRANSFORM, Shift Values, Visual Binning, Optimal Binning în SPSS

Programul SPSS are opțiuni în plus față de PSPP, prima dintre acestea fiind **Transform**, **Shift Values** care generează noi variabile prin modificarea seriei celor vechi. Cazurilor noii variabilei li se vor atribui valorile cazurilor anterioare sau următoare (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Graficul 7.24 Transform, Shift Values în SPSS



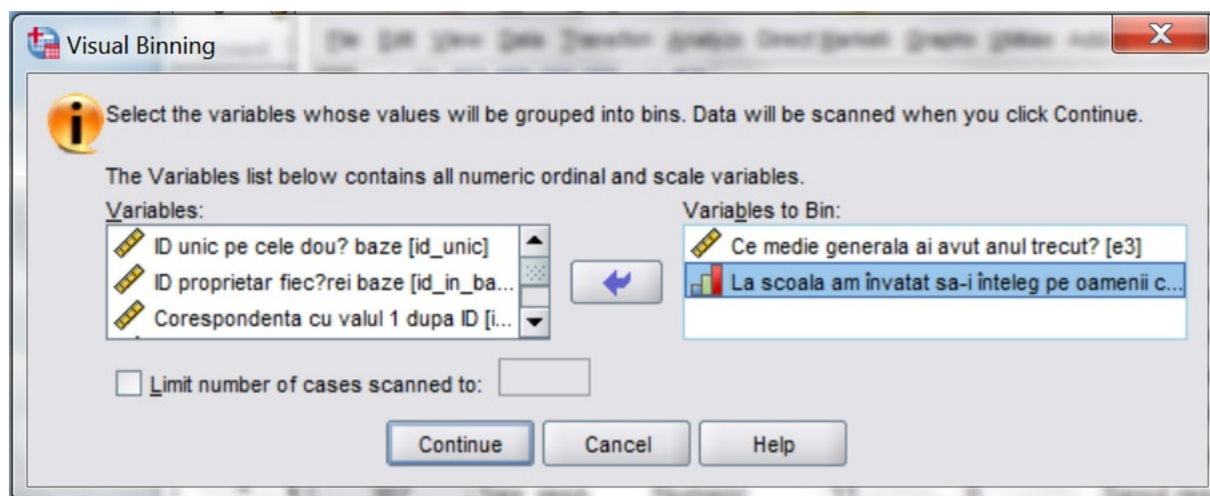
Dacă alegem „Get value from earlier case (Lag)”, fiecărui caz al noii variabile i se va atribui valoarea proprie cazului anterior al vechii variabile (primul caz va rămâne, în consecință, sub formă de dată lipsă).

Dacă alegem „Get value from later case (Lead)”, fiecărui caz al noii variabile i se va atribui valoarea proprie cazului următor al vechii variabile (ultimul caz va rămâne, deci, dată lipsă).

De notat că se poate stabili și cât să fie pasul (numărul cazurilor care trebuie „sărite” – „Number of cases to shift”). În exemplul din graficul 7.24 s-a trecut valoarea 1 dar putea fi stabilit orice alt număr întreg (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Opțiunea **Transform**, **Visual Binning** se regăsește doar în programul SPSS și are ca scop definirea de intervale de grupare pentru variabilele numerice cantitative și ordinale.

Grafic 7.25 Transform, Visual Binning în SPSS



Fereastra de dialog deschisă prin comenzile „Transform”, „Visual Binning”, din graficul 7.25, permite selectarea variabilelor (cantitative sau calitative ordinale) din lista aflată în partea stângă (Variables) cu scopul de a putea fi definite noile variabile cu valori grupate în intervale (Variables to Bin).

În partea de jos a graficului 7.25 există o opțiune de limitare a numărului de cazuri care sunt scanate (având ca scop cunoașterea caracteristicilor), opțiune care este utilă pentru fișierele foarte mari dar care e bine să fie evitată deoarece, când sunt luate în calcul toate cazurile dintr-o bază de date, e necesară întreaga informație pentru reușita procedurii (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

După ce s-au selectat variabilele pentru care se vor defini intervalele de grupare, se dă comanda „Continue” și se generează o nouă fereastră de dialog, ilustrată în graficul 7.26.

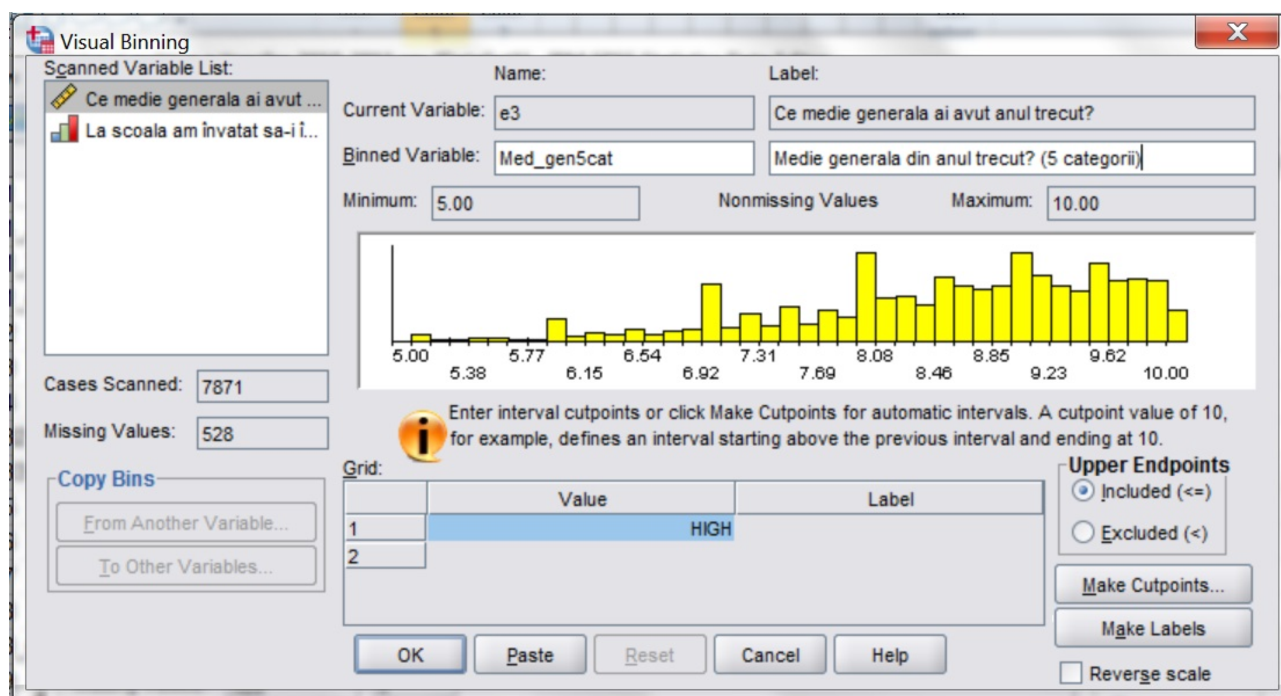
În partea stângă (sus) a ferestrei de dialog sunt listate variabilele care au fost selectate pentru construcția noilor variabile categoriale.

În stânga jos e prezentat numărul cazurilor scanate și cel al nonrăspunsurilor pentru variabila selectată în exemplul nostru având numele „e3” (Name) și eticheta „Ce medie generală ai avut anultrecut?” - label (în centru și dreapta sus).

Sub numele și eticheta variabilei care a fost selectată apar aceleași informații privind noua variabilă categorială, adică noile „name” și „label” (IBM SPSS Statistics for Windows, Version 23: A Basic Tutorial).

În partea centrală a ferestrei de dialog este prezentată histograma proprie cazurilor valide ale variabilei desemnând media generală. Valorile minime și maxime (în cazul nostru 5 și 10) sunt cele identificate de SPSS abstracție făcând de datele declarate lipsă.

Grafic 7.26 Transform, Visual Binning 2 în SPSS



Următorul pas al procedurii constă în definirea punctelor limită (cutpoints) folosind modalitățile oferite de SPSS („Intervale de lățime egală", „Percentile egale" și "Puncte de limită la medie și abaterile standard".

Graficul 7.27 Transform Visual Binning Make Cutpoints în SPSS

Make Cutpoints
Equal Width Intervals
Intervals - fill in at least two fields
First Cutpoint Location: 5.00
Number of Cutpoints: 5
Width: 1.000
Last Cutpoint Location: 9.00
Equal Percentiles Based on Scanned Cases
Intervals - fill in either field
Number of Cutpoints:
Width(%):
Cutpoints at Mean and Selected Standard Deviations Based on Scanned Cases
+/- 1 Std. Deviation
+/- 2 Std. Deviation
+/- 3 Std. Deviation
Apply will replace the current cutpoint definitions with this specification.
A final interval will include all remaining values: N cutpoints produce N+1 intervals.
Apply Cancel Help

Prima variantă (Equal Width intervals) presupune ca, mai întâi, să introducem primul punct limită („First cutpoint"). Am ales ca acest punct să fie limita inferioară a intervalului de variație al variabilei, adică valoarea 5. În continuare, trebuie să introducem numărul punctelor limită („Number of cutpoints"), fapt care va determina și lățimea intervalelor de grupare. În cazul nostru, evident, programul SPSS a stabilit lățimea egală cu 1 (**5**-5.99, **6**-6.99, **7**-7.99, **8**-8.99 și **9**-10). Ultimul punct de limită (Last Cutpoint Location) e indicat ca fiind 9 (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Dacă am fi optat pentru „Equal percentiles based on scanned cases" am fi definit numărul de puncte de limită („cutpoints") astfel încât acestea să delimiteze numere egale de cazuri (lățimea intervalului – width – este exprimată ca procent de cazuri propriu fiecărei categorii a noii variabile).

Ultima opțiune, „Cutpoints at mean and selected Standard Deviation Based on Scanned Cases" presupune, din nou, că punctele de limită („cutpoints") se vor afla la distanțe egale dar în acest caz valoarea de referință va fi media aritmetică iar distanțele la care vor fi plasate respectivele „cutpoints" pot să fie:

- +/– 1 abatere standard
- +/– 2 abateri standard
- +/– 3 abateri standard
- +/– 1 abatere standard și +/– 2 abateri standard
- +/– 1 abatere standard și +/– 3 abateri standard
- +/– 2 abateri standard și +/– 3 abateri standard
- +/– 1 abatere standard și +/– 2 abateri standard și +/– 3 abateri standard

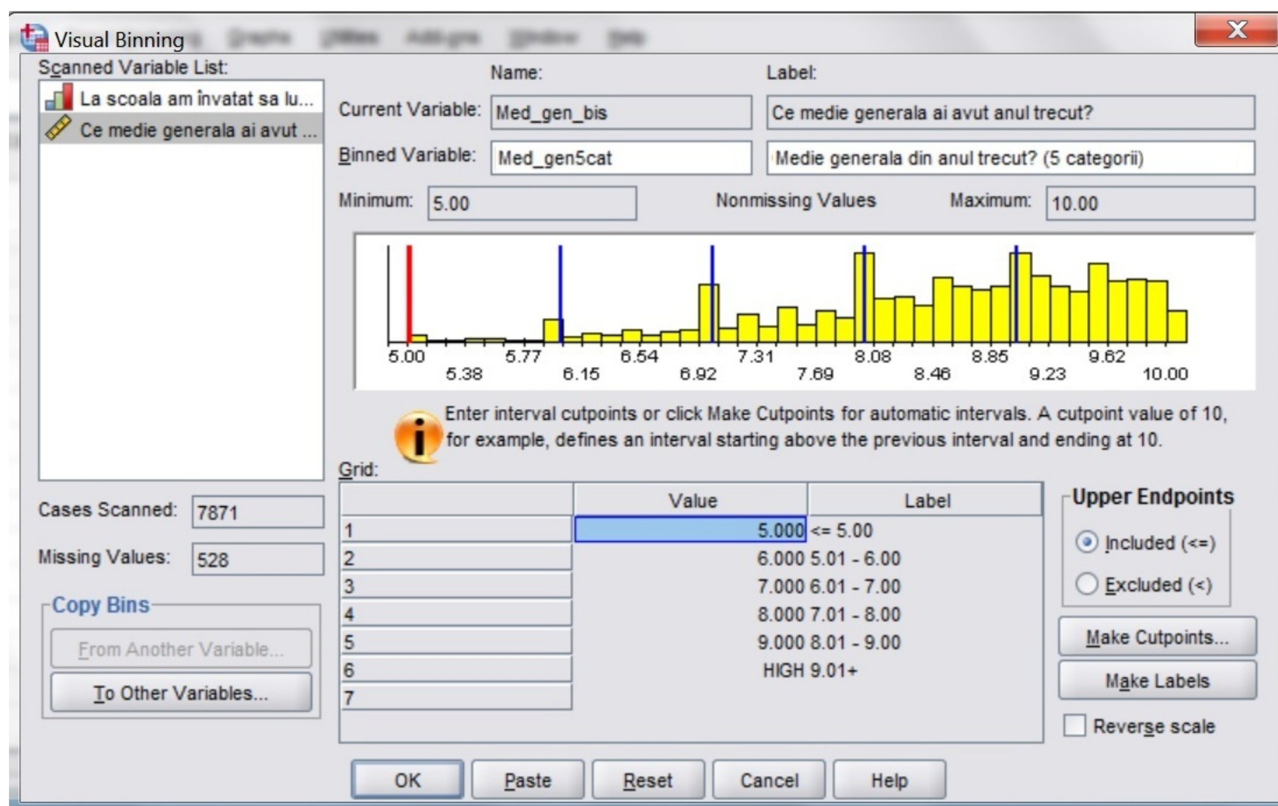
Ultima precizare din graficul 7.27 atrage atenția asupra faptului că acționând butonul „Apply" vom înlocui „cutpoints" care au fost setate anterior. În plus, se menționează că un interval final va include toate valorile necuprinse în primele N intervale definite de cele N „cutpoints" (ținând seama că N „cutpoints" definesc, de fapt, N+1 intervale).

Dacă, în continuare, în fereastra ilustrată în graficul 7.27 dăm comanda „Apply", vom obține o imagine asemănătoare cu cea din graficul 7.26 dar, de data aceasta, vor apărea valorile pentru „cutpoints" aliniate sub titulatura „Value" (în partea de jos a graficului). În exemplul nostru valorile pentru „cutpoints" sunt 5, 6, 7, 8 și 9 care definesc 6 intervale (ultimul fiind reprezentat de valorile mai mari sau egale cu 9.01).

Pentr a releva valorile subsumate fiecărui „cutpoint" e necesar să dăm comanda „Make labels" care determină afișarea intervalelor care sunt proprii „cutpoint"-urilor care au fost definite. Mai menționăm faptul că prin căsuța „Upper Endpoints" se poate stabili dacă valoarea exactă a „cutpoint"-ului să fie inclusă în propriul interval delimitat (Included ≤) sau să fie exclusă din acest interval (Excluded <).

În exemplul nostru, primul „cutpoint" va subsuma toate valorile variabilei mai mici sau egale cu 5 (în speță, toate mediile exct egale cu 5). Lui 6 îi vor corespunde valorile între 5.01 și 6, etc. Ultimului „cutpoint", care are valoarea 9, îi va corespunde intervalul 8.01-9 iar al șaseșea interval va fi definit prin valorile mai mari sau egale cu 9.01 (High 9.01). În cazul nostru, maximul va fi 10 (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 7.28 Transform Visual Binning Make Cutpoints în SPSS

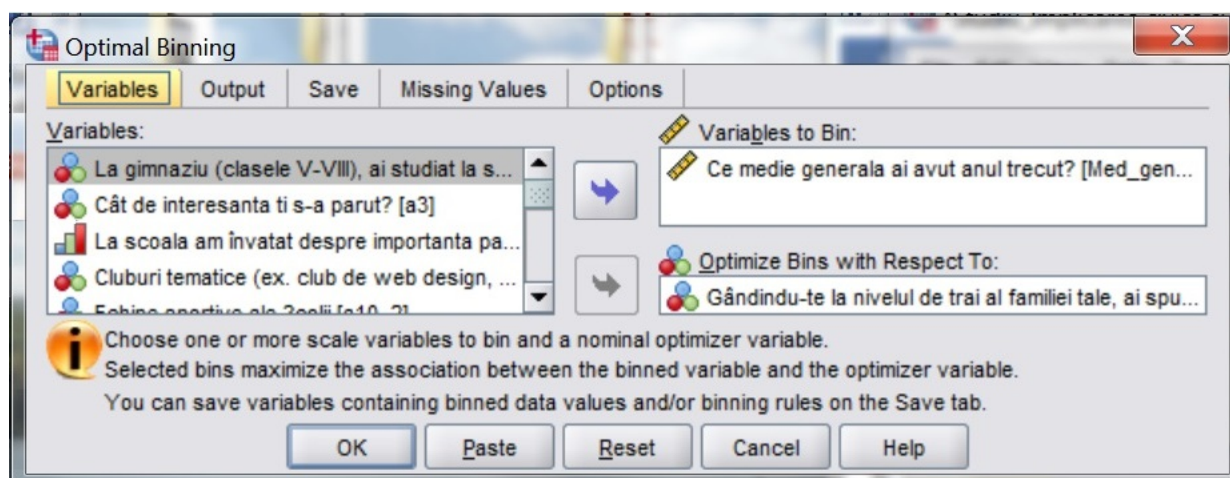


De notat că prin opțiunea „Copy Bins" „To other variables" se poate aplica aceeași procedură de atribuire a „cutpoint"-urilor și intervalelor corespunzătoare pentru o altă variabilă („To Other Variables") dacă, evident, au naturi similare (ar fi putut fi folosită media obținută în semestrul I sau II).

O altă facilitate a „Visual Binning" e faptul că barele care marchează „cutpoint"-urile pot fi deplasate în interiorul histogramei (în condițiile în care intervalele aferente rămân neschimbate).

Următoarea opțiune a meniului „Transform" o constituie „Optimal Binning" care transformămă în valori discrete una sau mai multe variabile cantitative continue distribuind valorile fiecărei variabile în intervale - „Bins" (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 7.29 Transform Optimal Binning în SPSS



Anterior ne-am ocupat de opțiunea „Visual Binning” care oferă o serie de metode automate de creare a intervalelor de grupare („Bins”) fără a utiliza o variabilă de ghidare. Aceste proceduri sunt utile pentru a produce statistici descriptive dar „Optimal Binning” tratează datele într-un mod mai adecvat dacă se urmărește crearea unui model predictiv (graficul 7.29).

Formarea intervalelor este optimizată în raport cu o variabilă de ghidare categorică care "supervizează" procesul de definire a categoriilor pentru fiecare variabilă de intrare. Astfel, gruparea pe intervale poate fi folosită în locul valorilor inițiale ale datelor pentru viitoarele analize (atunci când e preferabil să se opereze cu variabile categoriale, nu cu unele cantitative, cum e cazul tabelelor de asociere). În exemplul din graficul 7.29 am ales media generală de anul trecut ca variabilă care urmeză a fi supusă procedurii „Optimal Binning” iar percepția asupra nivelului de trai al familiei a fost selectată drept caracteristică categorială care optimizează procedura - „Optimize Bins with Respect To” (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Procedura se lansează prin comanda „Tramsform” „Optimal Binning” și apoi se selectează una sau mai multe variabile de intrare (cantitative) și o variabilă de ghidare (categorială, „string sau numerică”).

Ca date de ieșire, procedura generează tabele cu „cutpoints” pentru intervalele de grupare și statistici descriptive pentru fiecare variabilă cantitativă de intrare (graficul 7.30). Pentru fiecare variabilă cantitativă de intrare, se afișează numărul de cazuri valide și a celor lipsă, numărul de valori valide distincte și minimul / maximul. Pentru variabila de ghidare asociată fiecărei variabile de intrare e afișată distribuția categoriilor acesteia.

Graficul 7.30 Transform Optimal Binning (Output) în SPSS



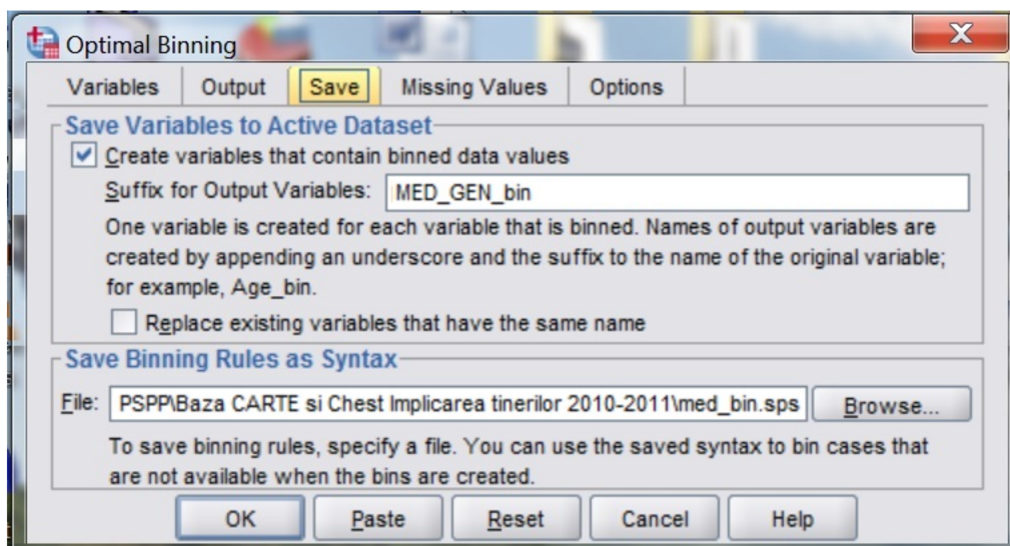
Ultima opțiune pentru datele de ieșire este „Model entropy for variables that are binned”. Pentru fiecare variabilă cantitativă de intrare, se afișează o măsură a preciziei predictive a acesteia în raport cu variabila de ghidare (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Opțiunile de salvare pentru procedura „Transform” „Optimal Binning” sunt prezentate în graficul 7.31. Se pot salva variabilele nou create în baza de date activă conținând „cutpoints” și intervalele care le sunt asociate.

În plus, putem salva sintaxa cu comanzile SPSS de transformare a variabilelor pentru a o utiliza în construcția intervalelor de grupare a unor noi variabile (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

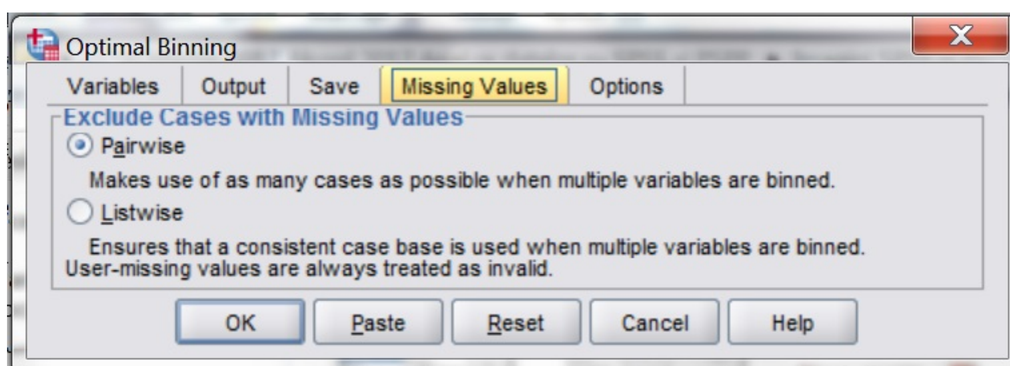
În graficul 7.31 putem vedea că variabilei categoriale nou-construită prin procedura „Optimal Binning” i s-a atribuit numele „MED_GEN_bin în conformitate cu cerința ca noile variabile să aibă sufixul „bin”. Regulile de sintaxă au fost salvate în fișierul med_bin.sps (Save Binning Rules as Syntax).

Graficul 7.31 Transform Optimal Binning (Save) în SPSS



În graficul 7.32 este prezentat modul în care sunt gestionate datele lipsă pentru procedura „Transform" „Optimal Binning". În primul rând, cu ajutorul căsuței de dialog „Save Variables to Active Dataset" putem stabili dacă datele lipsă sunt eliminate „Pairwise" (pereche) sau „Listwise" – listă (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 7.32 Transform Optimal Binning (Missing values) în SPSS



Este important de remarcat că datele definite ca fiind lipsă de către utilizator („user missing") vor fi întotdeauna tratate drept date lipsă. Atunci când se recodează valorile variabilelor inițiale într-unele noi, datele lipsă definite de utilizator sunt convertite în „missing system" (informații lipsă în baza de date).

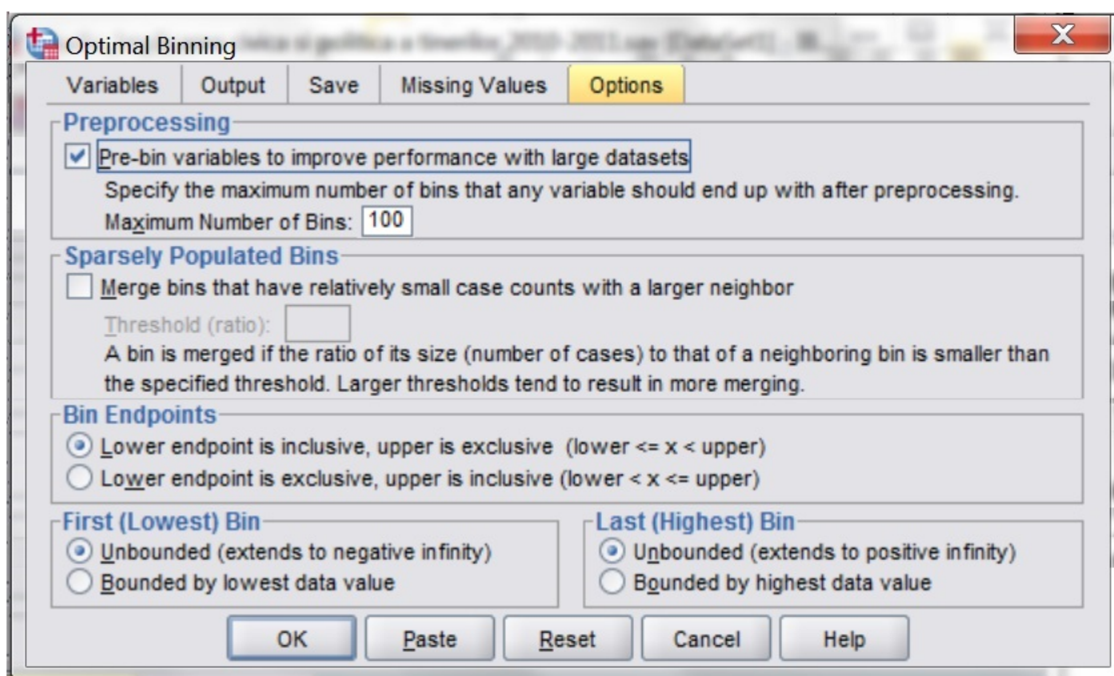
Opțiunea de excludere a cazurilor „Pairwise" (pereche) vizează fiecare pereche formată dintr-o variabilă de intrare și cea de ghidare. Procedura va utiliza toate cazurile cu valori valide disponibile atât pentru variabila de ghidare cât și pentru cea cantitativă, de intrare.

Opțiunea de excludere a cazurilor „Listwise" vizează toate variabilele specificate în căsuța „Variabiles to bin". Dacă o valoare a uneia dintre variabile lipsește pentru un anumit caz, întregul caz respectiv este exclus. Altfel spus, nu vor fi luate în considerare valorile valide pentru alte variabile cantitative de intrare (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Ultima dintre fereastrele care pot fi lansate prin comenzile „Transform", „Optimal Binning" este „Options".

În graficul 7.33 sunt ilustrate posibilitățile ferestrei de dialog „Options". Prima chestiune se referă la posibilitatea pre-procesării variabilelor de intrare care au număr mare de valori distincte fără a fi afectată semnificativ calitatea finală a recodificării acestora (recodificare ghidată de variabila categorială aleasă).

Graficul 7.33 Transform Optimal Binning (Options) în SPSS



În căsuța „Preprocessing" putem specifica numărul maxim de intervale care vor fi create prin procesul de pre-recodificare a variabilelor cantitative de intrare (Maximum number of bins). De notat că dacă se specifică 100 ca valoare maximă dar o variabilă cantitativă de intrare are mai puțin de 100 de valori distincte, numărul de intervale pre-procesate create pentru respectiva variabila de intrare va fi egal cu numărul de valori distincte al acesteia (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

În căsuța „Sparsely Populated Bins" din „Options" se referă la situația în care prin recodificarea variabilelor cantitative se obțin intervale cu foarte puține cazuri. În această situație vom selecta „Merge bins that have relatively small case counts with a larger neighbor". Pentru o anumită variabilă, să presupunem că algoritmul a găsit N_{final} „cutpoints" cărora le corespund N_{final} +1 intervale. Pentru intervalele i = 2, 3, ... N_{final} (între a doua cea mai mică valoare a intervalelor și a doua cea mai mare) se calculează:

Mărimea lui f(b_i) / min [mărimea lui (b_{i-1}), mărimea lui (b_{i+1})],
unde mărimea lui (b) reprezintă numărul de cazuri din interval

Când mărimea calculată în formulă este mai mică decât decât pragul de fuzionare specificat, b_i e considerat slab populat și e contopit cu b_{i-1} sau b_{i+1}, oricare dintre acestea care are o entropia informațională de intrval mai mică (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Opțiunea „Bin Endpoints" din „Options" ne permite să alegem dacă intervalele generate vor cuprinde marginea din stânga (lower ≤ x < upper) sau marginea din dreapta (lower < x ≤ upper).

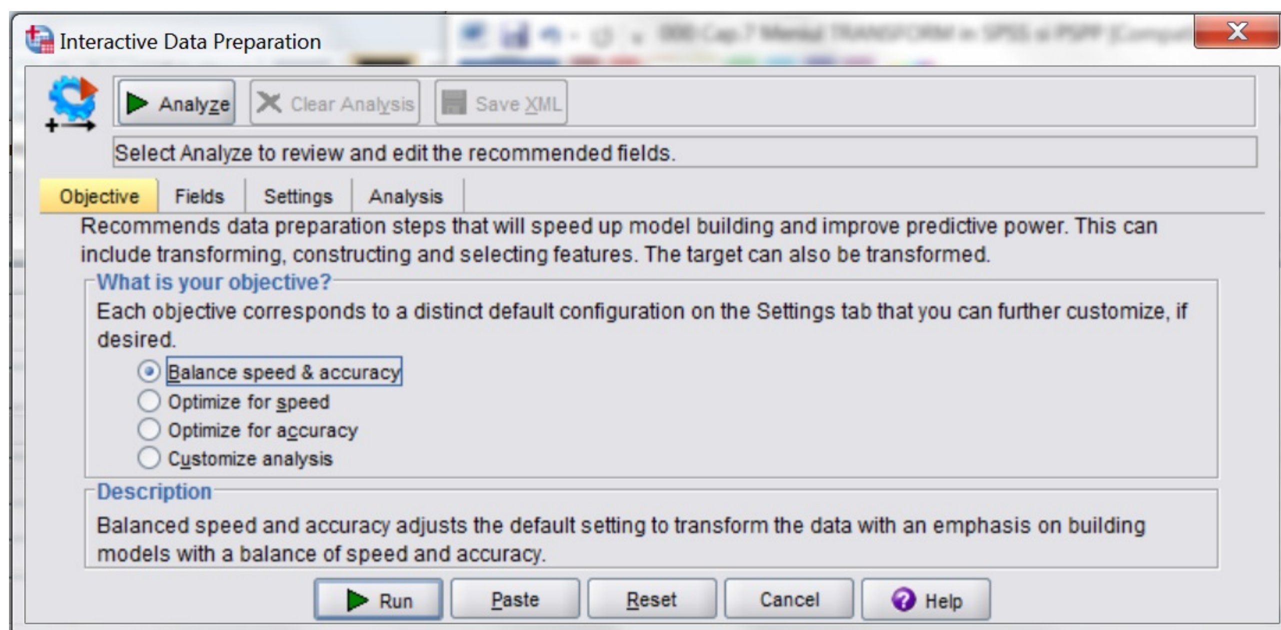
Opțiunea „Optimal binning, options” (graficul 7.31) vizează limitele primului și ultimului interval obținut recodificând variabilele cantitative de intrare. Alegând „Unbounded” pentru „First (lowes) Bin” (primul interval), limita stângă a intervalului va fi minus infinit. Analog, alegând „Unbounded” pentru „Last (highest) Bin” (ultimul interval), limita din dreapta a intervalului va fi plus infinit. Uzual se aleg limitele definite de către valorile minime și maxime ale variabilei cantitative de intrare - „Bounded by lowest / highest data value” (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

7.5 Opțiunile TRANSFORM, Prepare Data For Modeling în SPSS

O altă opțiune a meniului **Transform** este **Prepare Data For Modeling** și vizează pregătirea datelor pentru analiză. Pregătirea automată a datelor presupune analiza informațiilor și identificarea corecțiilor necesare, examinearea domeniilor problematice, obținerea de noi variabile (atunci când este nevoie) și îmbunătățirea performanței globale prin tehnici inteligente de scanare a informațiilor. Algoritmul poate fi utilizat complet automat, dându-i-se posibilitatea să facă alegeri și să aplice corecții sau poate fi folosit interactiv, utilizatorul putând cunoaște și accepta sau respine modificările înainte de a fi realizate în cadrul procedurii.

Vom începe prezentarea opțiunii „Transform”, „Prepare Data For Modeling” referindu-ne la varianta „Interactive Data Preparation”. Prima fereastră de dialog este „Objective” și presupune alegerea între 4 variante (graficul 7.34).

Graficul 7.34 Transform, Interactive Data Preparation, Objective în SPSS



Pentru a identifica diferențele dintre cele patru variante de obiective, care corespund unor configurații implicite distincte ale setărilor, le vom compara lansându-le în mod succesiv.

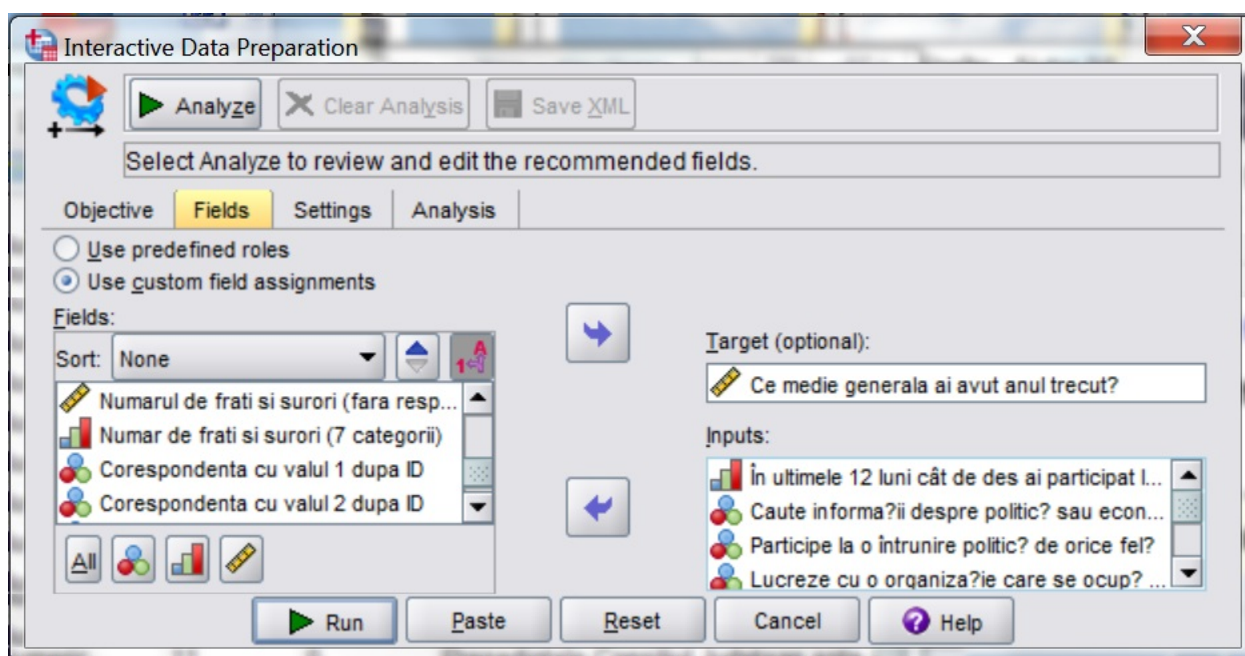
Vom începe cu varianta „Balance speed & accuracy”. Înainte de a lansa procedura „Analyze” din partea de sus a ferestrei (ilustrată în graficul 7.34), vom da comanda „Fields” care controlează ce variabile vor fi procesate pentru viitoarele analize (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Prima opțiune a ferestrei ne permite să alegem între a utiliza rolurile predefinite („Use predefined roles") și a face modificări în respectivele roluri (Use custom field assignments).

Dacă alegem „Use predefined roles", această variantă va folosi informațiile existente despre variabile. Dacă există o singură variabilă care are rolul de țintă („tareget"), acesta va fi utilizată ca variabilă dependentă; în caz contrar, nici o variabil nu va fi aleasă ca țintă. Toate variabilele cu rol de intrare (input) predefinit vor fi utilizate ca atare. Procedura necesită cel puțin o variabilă de intrare.

Menționăm faptul ca în stânga ferestrei se găsește lista completă de variabile din baza de date pe când în partea dreapta sunt afișate cele care au fost selectate automat ca variabile de intrare („inputs"). Nici un schimb nu este posibil intre cele două liste și nici alegerea unei variabile „target" (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 7.35 Transform, Interactive Data Preparation, Fields în SPSS



Dacă alegem „Use custom field assignments" vom avea posibilitatea de a elimina variabile input sau de a introduce altele noi din întregul câmp de caracteristici definite în baza de date. În plus, va putea fi aleasă variabila de tip „Target" pentru viitorul model explicativ (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

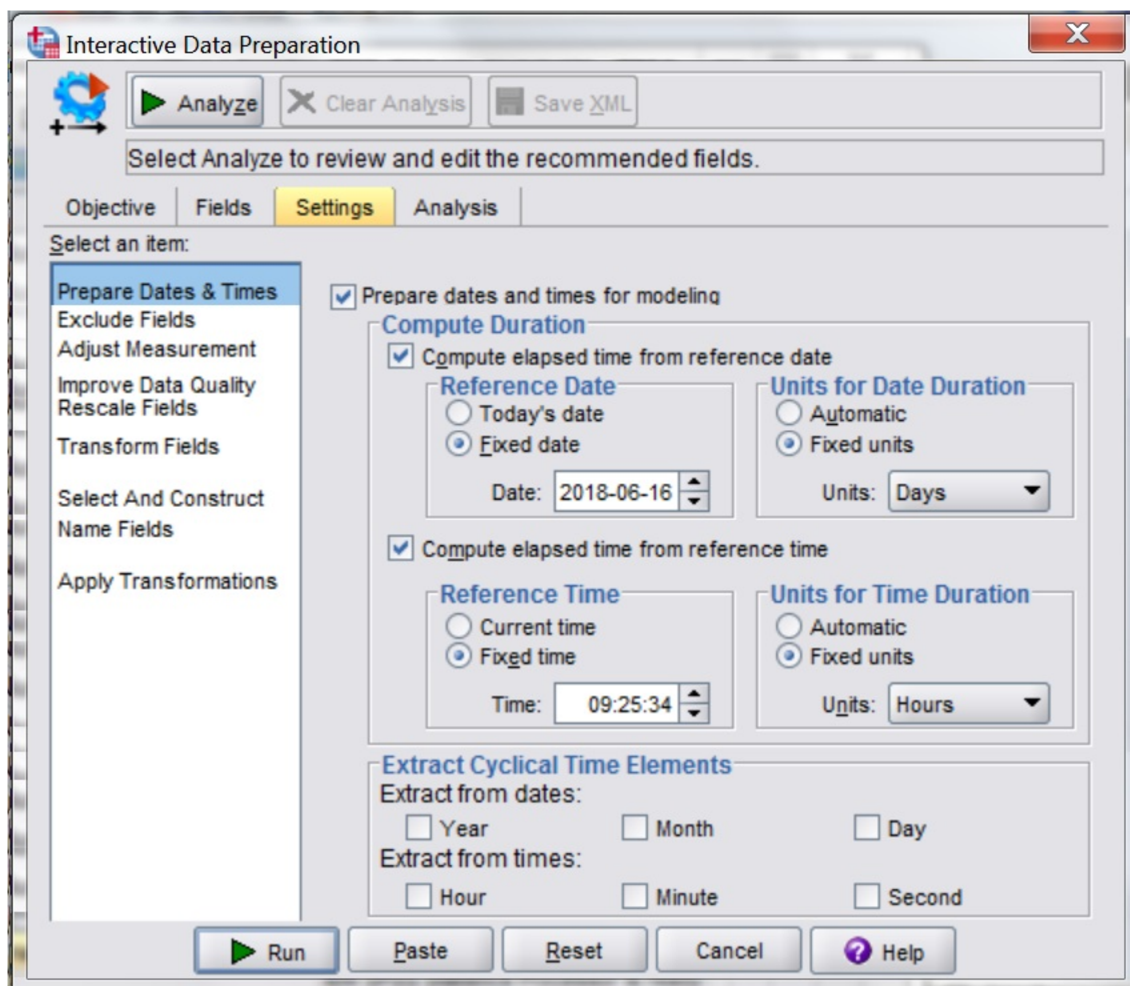
Următoarea fereastră a procedurii „Interactive data preparation" este „Settings" care cuprinde mai multe grupuri diferite de setări care pot fi modificate pentru a regla modul de procesare a datelor de către algoritmul programului SPSS.

Dacă se efectuează modificări ale setărilor implicite care sunt incompatibile cu celelalte obiective, fereastra „Objective" este actualizată automat pentru a selecta opțiunile personlizate („Customize analysis").

Prima setare, „prepare dates & time", vizează timpul și data (gaficul 7.36) și are rolul de a pregăti respectivele dimensiuni pentru modelare. Opțiunea „Compute duration" are două aspecte distincte, măsurarea timpului rămas de la o dată de referință respectiv a timpului rămas de la un moment de timp (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

În fiecare dintre cele două variante se poate alege reperul prezent (data respectiv momentul de timp prezent) sau se pot fixa anumite repere („Fixed date" și Fixed time"). În plus, atât pentru calcularea timpului de la o dată de referință cât și în raport cu un moment de timp se pot fixa automat – „Automatic" – unitățile de măsură sau pot fi stabilite de utilizator – „Fixed units" (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 7.36 Transform, Interactive Data Preparation, Settings în SPSS



A doua setare, „Exclude fields", se referă la excluderea datelor de intrare de slabă calitate. Pot fi excluse, în primul rând, variabilele cu prea multe date lipsă (se stabilește un procentaj maxim admisibil al datelor lipsă).

Există, de asemenea, posibilitatea excluderii variabilelor nominale cu un număr prea mare de categorii diferite (este setat un număr maxim de categorii). Ultima posibilitate de excludere vizează variabilele categoriale cu prea multe valori într-o singură categorie (se stabilește procentajul maxim / categorie). În ultima parte a ferestrei „Exclude fields" ni se atrage atenția că acele variabile cu valori constante vor fi întotdeauna excluse.

A treia setare „Adjust mesurement" se referă la numărul maxim de valori pentru variabilele ordinale („Maximum number of values for ordinal fields") și la numărul minim de valori distincte pentru variabielele cantitative („Minimum number of values for continuous fields". Menționăm faptul că se poate opta ca limitările să fie aplicate variabilelor „Inputs" și / sau „Target" IBM Knowledge Center, Home, SPSS Statistics 24.0.0.

A patra opțiune din fereastra „Settings" se referă la pregătirea seturilor de variabile pentru îmbunătățirea caliății datelor („Prepare fields to improve data quality"). Prima opțiune de creștere a calității datelor se referă la gestionarea valorilor „Outlier" (valori izolate, care diferă foarte mult de medie). În primul rând se poate alege înlocuirea valorilor „Outlier" pentru variabilele „Inputs" și / sau pentru „Target". În al dolea rând se poate defini care sunt valorile considerate „Outlier" indicând numărul de abateri standard de la medie de la care datele nu vor mai fi acceptate. Se propune, în continuare, o metodă de tratare a „Outlier" – ilor: înlocuirea cu valoarea corespunzăroare abaterilor standard stabilite anterior ca fiind maximumul acceptabil sau definirea acestora ca date lipsă. O nouă căsuță de dialog propune variante de înlocuire a datelor lipsă (Replace Missing Values), care se referă la variabilele de intrare și / sau la „Target":

- Variabilele nominale: înlocuirea datelor lipsă cu valoarea modală
- Variabile ordinale: înlocuirea datelor lipsă cu valoarea mediană
- Variabile cantitative continue înlocuirea datelor lipsă cu media (IBM Knowledge Center, Home, SPSS Statistics 24.0.0)

Ultima chestiune legată de pregătirea variabilelor pentru îmbunătățirea calității datelor constă în „Reorder Nominal Fields" și se referă la variabilele „Inputs" și / sau „Target". Putem opta pentru a reordona variabilele nominale astfel încât prima categorie să aibă frecvența cea mai mică iar ultima, cea mai mare.

Dând comanda „Analize" procedura inițiată se va centra automat asupra ferestrei „Analysis" în care sunt procesate datele.

A cincea posibilitate oferită de fereastra „Settings" se referă la refacerea scalării variabilelor („Rescale fields"). Prima căsuță de dialog este „Analysis Weight". Aici trebuie introdusă o variabilă cantitativă continuă care conține ponderări pentru analiză (regresie sau eșantionare). Următoarea căsuță de dialog e denumită „Continuous Input Fields" și are ca efect normalizarea variabilelor „input" continue utilizând o transformare a scorului z sau o transformare min / max. Rescalarea variabilelor de intrare este utilă mai ales atunci când se selectează „Perform feature construction" în varianta de setare „Select and Construct". Notăm faptul că pentru transformarea scorului „Z" putem stabili valoarea finală a mediei (implicit fiind 0) și cea finală a abaterii standard (implicit 1). Dacă ne folosim de tansformarea min / max, se poate stabili valoarea minimă (implicit 0) și cea maximă (implicit 100).

Ultima căsuță de dialog pentru „Rescale Fields" este „Continuous Target" care transformă o variabilă țintă, prin intermediul unei transformări Box-Cox, într-o variabilă cu o distribuție aproximativ normală, având specificate abaterea standard finală și media finală. Se pot alege valoarea finală a mediei și un număr pozitiv pentru valoarea abaterii standard finale. În mod implicit, valorile anterior amintite vor fi 0 și, respectiv, 1 (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

A șasea opțiune din fereastra „Interactive Data Preparation" „Settings" vizează creșterea puterii predictive a datelor prin transformarea variabilelor „input". Butonul „Transform field for modeling" trebuie selectat pentru a fi active opțiunile de control al variabilelor. Casuța de dialog „Categorical Input Fields" începe cu „Merge sparse categories to maximize association with target". Rolul acesteia e de a crea un model mai simplu, prin reducerea numărului de categorii care vor fi procesate în asociere cu variabila „target".

Categoriile similare sunt identificate pe baza relației dintre variabilele de intrare și „target". Categoriile care nu sunt semnificativ diferite (adică, care au o valoare „P" mai mare decât valoarea specificată) sunt sumate.

Trebuie specificată o valoare mai mare decât 0 și cel mult egală cu 1. Dacă toate categoriile sunt sumate într-una singură, versiunea inițială și cea derivată a variabilei sunt excluse din analiza ulterioară deoarece nu au valoare ca predictor. În lipsa variabilei țintă, există opțiunea „When there is no target, merge sparse categories based on counts" care permite sumarea categoriilor variabilelor în funcție de frecvențele acestora pentru caracteristicile ordinale și nominale (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Metoda frecvenței egale este folosită pentru a însuma categoriile care dețin un procent mai mic de cazuri din total decât cel minim specificat. Se va indica o valoare pozitivă, mai mică ori egală cu 100. Valoarea implicită e 10. Sumarea se oprește atunci când nu mai există categorii cu un procent de cazuri inferior celui minim specificat sau când rămân doar două categorii.

Ultima opțiune din „Transform fields" este „Continuous Input Fields" și e utilizată dacă variabila țintă e categorială pentru a introduce în analiză variabile de intrare cantitative puternic asociate cu aceasta (transformate, evident, în variabile categoriale).

Categoriile sunt create pe baza proprietăților „subseturilor omogene", care sunt identificate prin metoda Scheffe de determinare a subseturile omogene (utilizând valoarea „P" specificată drept „alfa" pentru valoarea critică).

Trebuie indicată o valoare mai mare decât 0 și mai mică sau egală cu 1. Valoarea implicită este 0,05. Dacă operația de construcție a categoriilor are ca rezultat un singur interval pentru o anumită variabilă, variantele originală și categorială ale variabilei sunt excluse deoarece nu au valoare predictivă (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Trebuie să ținem cont de următorul fapt: construirea intervalelor de grupare „automată" diferă de același proces așa numit „optimal". „Binning-ul" optimal folosește informații despre entropie pentru a converti o variabilă continuă într-una categorială; acest lucru presupune sortarea datelor și stocarea lor în memorie.

Construirea automată a categoriilor folosește subseturi omogene pentru a împărți pe intervale o variabilă continuă, ceea ce înseamnă că pregătirea „binning-ului" automat nu necesită sortarea datelor și stocarea tuturora în memorie.

Utilizarea metodei subsetului omogen în operația de construcție a categoriilor pentru variabilele continue conduce întotdeauna, după construirea intervalelor, la un numărul de categorii care este mai mic sau cel mult egal cu cel inițial.

A șaptea opțiune din fereastra „Interactive Data Preparation" „Settings" este intitulată „Select and Construct" și urmărește îmbunătățirea puterii predictive a datelor prin construirea de noi variabile pe baza celor existente.

Prima căsuță de dialog, intitulată „Feature selection", dacă e activată, va determina eliminarea din analiză a variabilelor cantitative continue dacă pragul de probabilitate al corelației cu variabila „target" e mai mare decât cel specificat în respectiva căsuță (se specifică în dreptul lui „p-value"). A doua căsuță de dialog se numește „Feature Construction" și are ca scop obținerea de noi variabile din combinarea mai multora existente în baza de date. Vechile variabile nu vor mai fi folosite în viitoarele analize. Această opțiune se aplică doar variabilelor de intrare cantitative continue cu variabilă „target" având aceeași natură sau când variabila „țintă" lipsește (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

A opta posibilitate oferită de fereastra „Interactive Data Preparation" „Settings" este „Name Fields" care urmărește identificarea cât mai ușoară a vriabilelor transformate și a celor noi, creând și aplicând, în mod automat, noi nume, prefixe sau sufixe.

Numele la care făceam referire pot fi modificate de către utilizator astfel încât să devină mai relevante pentru datele de care dispune și pentru trebuințele sale specifice (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Alegând „Transformed and Constructed Fields" vom putea stabili numele extensiei pentru variabilele de input transformate și pentru variabila țintă transformată (în mod implicit, numele extensiei va fi _transformed).

De asemenea, vom putea stabili numele prefixului aplicat oricărei variabile construite. Noul nume va fi creat atașând unui sufix numeric la acest nume de prefix.

Următoarea căsuță de dialog din „Name Fields" e intitulată „Durations Computed from Dates and Times" și permite specificarea extensiilor pentru numele care urmează să fie aplicate duratelor calculate atât pentru date (implicit: _years, _month și _dais) cât și pentru ore (implicit: _hours, _minutes și _seconds).

Ultima căsuță de dialog pentru „Name Fields" este „Extracted Cyclical Time Elements" care oferă posibilitatea specificării extensiilor de nume care pot fi aplicate elementelor ciclice extrase atât din cuantificarea datei cât și din cea a timpului.

A noua și ultima opțiune propusă de fereastra „Interactive Data Preparation" „Settings" este „Apply Transformation" și este utilă în aplicarea și salvarea transformărilor din procesul de pregătire automată a datelor.

Prima căsuță a ferestrei de dialog se intitulează „Transformed Data" și are rolul de a specifica unde vor fi salvate datele transformate.

Selectarea butonului „Add new fields to the active dataset" va face ca toate variabilele create prin procedura de pregătire automată a datelor să fie adăugate sub formă de noi variabile în baza de date care este activă.

Bifarea opțiunii „Update roles for analyzed fields" va actualiza situația tuturor variabilele care au fost excluse din analiza ulterioară prin intermediul pregătirii automate a datelor.

Următoarea opțiune este "Create a new dataset or file" și presupune generarea unei noi baze de date sau a unui fișier care conține datele transformate (sunt adăugate variabilele recomandate în urma procesului de pregătire automată a datelor).

Optând pentru „Include unanalyzed fields" vom adăuga variabile în baza de date date originală care nu au fost specificate în noa bază de date. Acest lucru este util deoarece permite transferul variabilelor care conțin informații neutilizate în modelare (în noul set de date) cum ar fi ID-ul, adresa sau numele.

Ultima opțiune pentru „Interactive Data Preparation" „Settings", „Apply Transformation", este „Location", care permite alegerea unui nume pentru baza de date nou generată (butonul „Dataset") sau salvarea acestuia indicând numele și calea către folderul de destinație (File).

Revenind la punctul de plecare al demersului nostru, compararea obiectivelor formulate în graficul 7.34 („Balance speed & accuracy", „Optimize for Speed" și „Optimize for accuracy"), alegem „Balance speed & accuracy" și apoi dăm comanda „Run".

Imaginea care apare implicit e „Field Processing Summary", care oferă o perspectivă de ansamblu asupra modului în care au fost procesate domeniile de analiză prin pregătirea automată a datelor (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Am stabilit o „țintă” (Target) folosind fereastra „Settings” reprezentată de media generală, sub formă cantitativă continuă. Au fost identificați 94 de predictori (variabile de intrare – input features) și un total de 94 de predictori dintre care 93 recomandați pentru utilizarea lor în analiză (construirea de modele). Dintre predictorii recomandați pentru utilizarea în analiză, 25 sunt variabile provenite din setul original de intrare (original fields) iar 68 se constituie într-un set de variabile transformate din predictorii inițiali. În analiză automată a datelor n-au fost recomandați predictori care să provină din variabile exprimând data / timpul sau construiți în procesul de pregătire automată a datelor (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

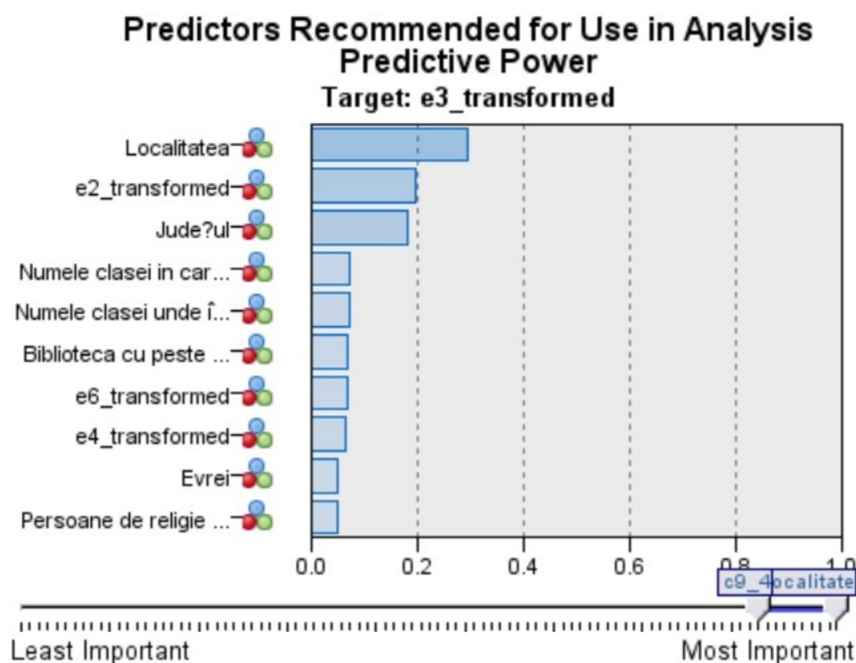
Graficul 7.37 Interactive Data Preparation, Balance speed & accuracy în SPSS

Field Processing Summary

Fields		N
Target		1
Predictors		94
Predictors recommended for use in analysis	Total	93
	Original fields (untransformed)	25
	Transformations of original fields	68
	Derived from dates and times	0
	Constructed	0
Predictors not used		1

A doua imagine oferită implicit de procedura în derulare e cea a puterii predictive („Predictive Power”) care oferă o imagine rapidă asupra variabilelor recomandate ca fiind cele mai utile pentru construirea de modele.

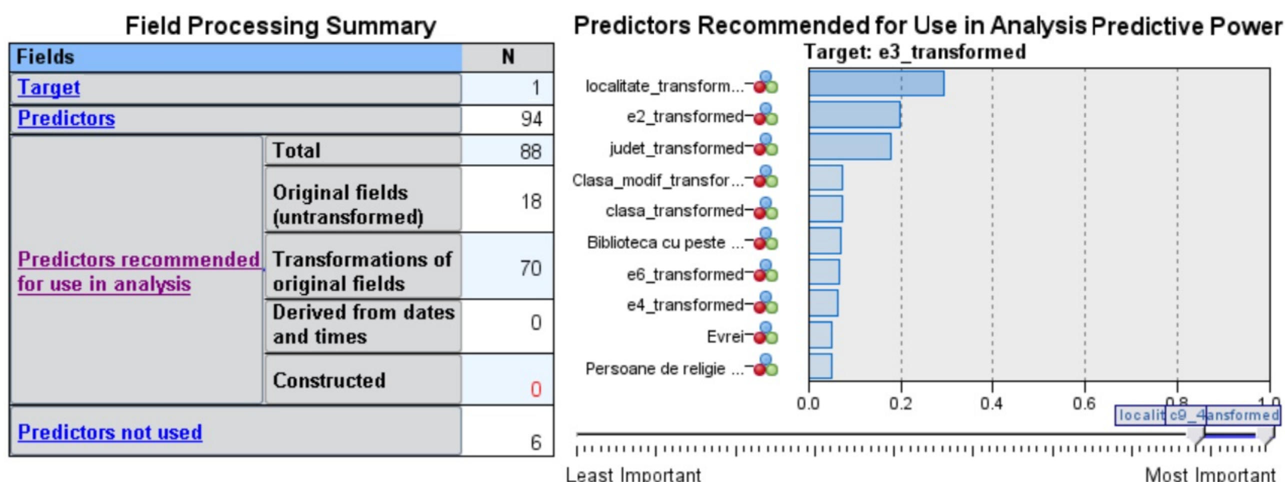
Graficul 7.38 Interactive Data Preparation, Balance speed & accuracy în SPSS



În graficul 7.37 se poate vedea că prin „Interactive Data Preparation” fixând obiectivul „Balance speed & accuracy”, primii 10 predictori cei mai importanți sunt variabile categoriale, localitatea de proveniență situându-se pe primul loc, urmată de variabila e2 transformată (faptul ca elevul să fi luat vreun premiu pentru activitatea școlară) și de județul de proveniență.

Alegând varianta „Optimize for Speed”, număru de predictori este același dar cei recomandați pentru a fi folosiți în analiză sunt doar 88. Variabilele păstrate în forma originală sunt doar 18 iar cele transformate sunt mai multe (70). Predictorii care nu au fost folosiți sunt 6, adică mai mulți cu 5.

Graficul 7.39 Interactive data preparation, Optimize for Speed in SPSS



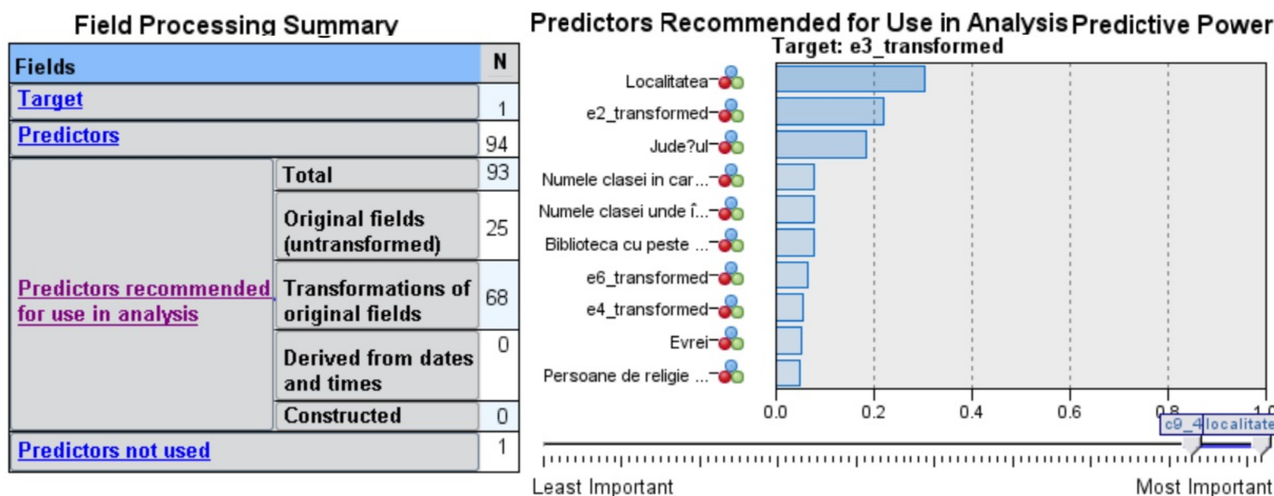
Field Processing Summary

Fields		N
Target		1
Predictors		94
Predictors recommended for use in analysis	Total	88
	Original fields (untransformed)	18
	Transformations of original fields	70
	Derived from dates and times	0
	Constructed	0
Predictors not used		6

Predictorii recomandați pentru a fi folosiți în modelul predictiv diferă, inclusiv cel dintâi (se propune „localitate_transformed” în locul variantei originale a variabilei). Al doilea predictor își păstrează forma generată de obiectivul „Balance speed & accuracy” iar al treilea este înlocuit cu varianta transformată a variabilei (judet_transformed”)

Alegând ca obiectiv „Optimize for Accuracy” avem o situație a alegerii și transformării predictorilor similară cu cea proprie alegerii obiectivului „Balance speed & accuracy” iar primele 10 variabile ca putere predictivă sunt identice (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 7.40 Interactive data preparation Optimize for Accuracy in SPSS

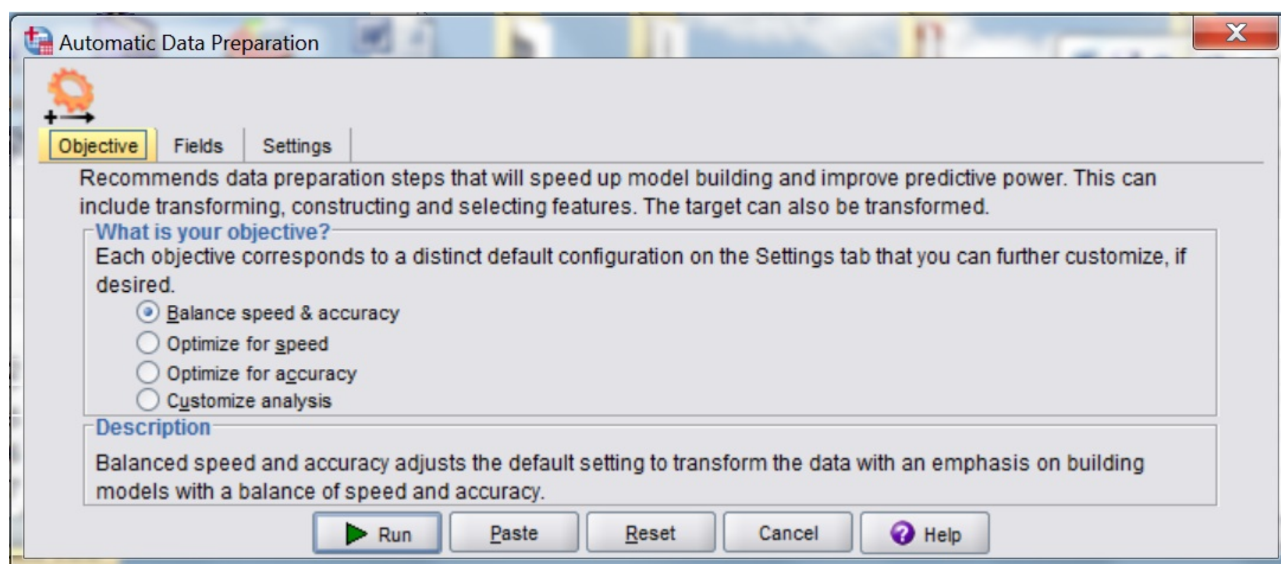


Field Processing Summary

Fields		N
Target		1
Predictors		94
Predictors recommended for use in analysis	Total	93
	Original fields (untransformed)	25
	Transformations of original fields	68
	Derived from dates and times	0
	Constructed	0
Predictors not used		1

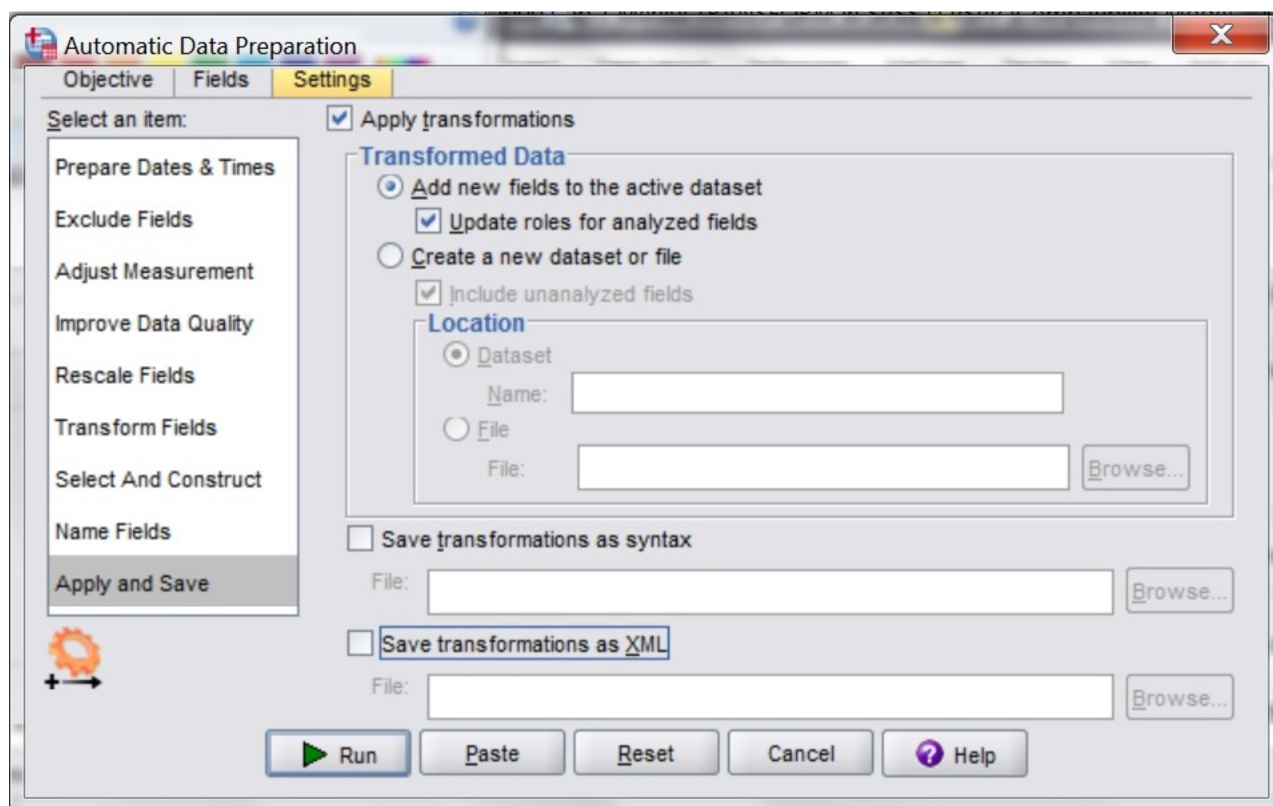
În continuare vom prezenta opțiunea „Transform”, „Prepare Data For Modeling” cu varianta „Automatic Data Preparation”.

Graficul 7.41 Transform, Automatic Data Preparation, Objective



Se poate observa că diferența față de fereastra „Interactiv data preparation” constă în lipsa opțiunilor „Analyze” (împreună cu „Clear Analysis” și Save XMLX). În rest, Graficul 7.41 e similar cu 7.34: trebuie optat, în primul rând, pentru un obiectiv care corespunde unei configurații implicite pentru fereastra „Settings”. Ulterior se pot face modificări dacă se dorește acest lucru (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

Graficul 7.42 Transform, Automatic Data Preparation, Settings

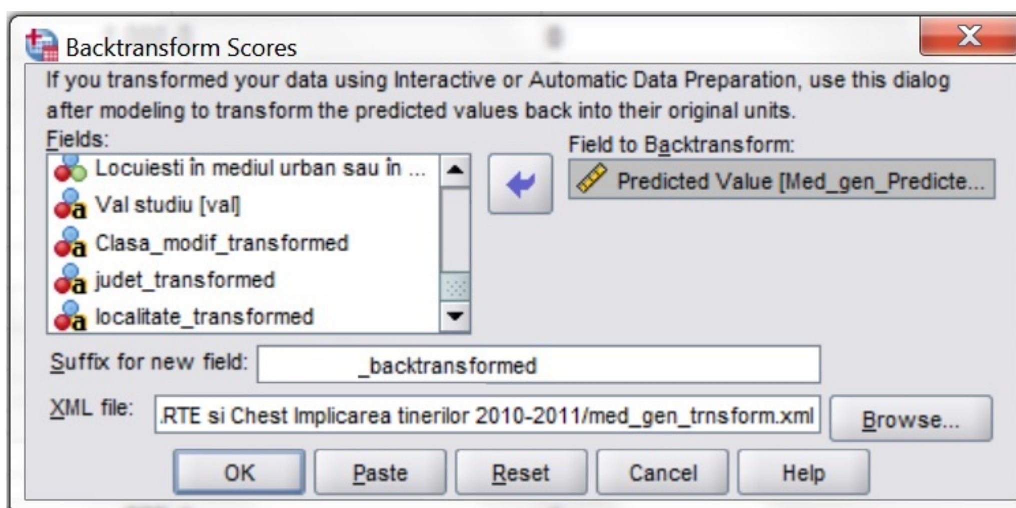


Ferestrele „Fields” și „Settings” din graficul 7.34 (Interactive data preparation) sunt asemănătoare cu cele din graficul 7.41 (Automatic Data Preparation) cu o diferență care apare la „Settings”, ultimele opțiuni: „Apply Transformation” – pentru „Interactive data preparation” respectiv „Apply and save” pentru „Automatic Data Preparation”. În graficul 7.42 este prezentată fereastra „Settings” („Automatic Data Preparation”) pentru opțiunea „Apply and Save” care oferă posibilitatea atribuirii unui nume și salvării noii baze de date construită prin procedura statistică aplicată. De asemenea, există oportunitatea salvării transformărilor aplicate variabilelor din baza de date sub formă de sintaxă conținând comenzile programului SPSS care au fost aplicate.

Ultima opțiune constă în salvarea transformărilor aplicate bazei de date inițială sub formă de fișier având extensia .xml și conținâd linii de program (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

A treia modalitate de pregătire automată a datelor este „Transform”, „Prepare Data for Modeling”, „Backtransform Scores”. Procedura e utilă dacă o variabilă țintă („target”) a fost transformată prin „pregătirea automată a datelor” și, prin urmare, modelele construite folosind acel „target” transformat generează scoruri specifice variabilelor transformate. Pentru interpretarea și utilizarea rezultatelor, trebuie convertită valoarea previzionată înapoi la scala inițială. Pentru a calcula scorurile „backtransform” trebuie dată comanda „Transform, Prepare Data for Modeling Backtransform Scores” iar în fereastra obținută trebuie selectată o variabilă care conține valori estimate de model (de exemplu prin analiză de regresie) ale caracteristicii „target” transformată (variabila trebuie să cuprindă valori estimate de model ale „țintei” transformate).

Graficul 7.43 Transform, Prepare Data for Modeling, Backtransform Scores



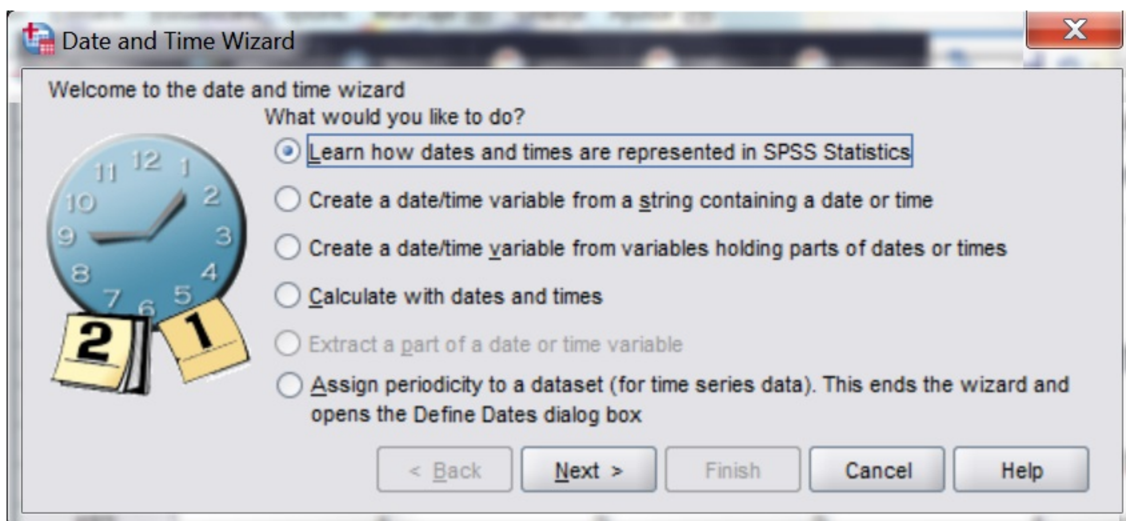
În continuare trebuie specificat sufixul pentru pentru noul câmp de variabile care va conține variabilele prezise (explicate) de către model dar scalate în forma originală, înainte de a fi supuse transformării.

Pasul care trebuie urmat ăn continuare îl constituie introducerea adresei în care se află fișierul cu extensia .xml și care conține transformările care au fost aplicate variabilelor. Acesta ar trebui să fie un fișier salvat din ferestrele „Interactive” sau „Automatic Data Preparation”. Dând comanda „Ok” vor fi obținute variabilele de interes în formatul inițial (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

7.6 Opțiunea TRANSFORM, Data and Time Wizard

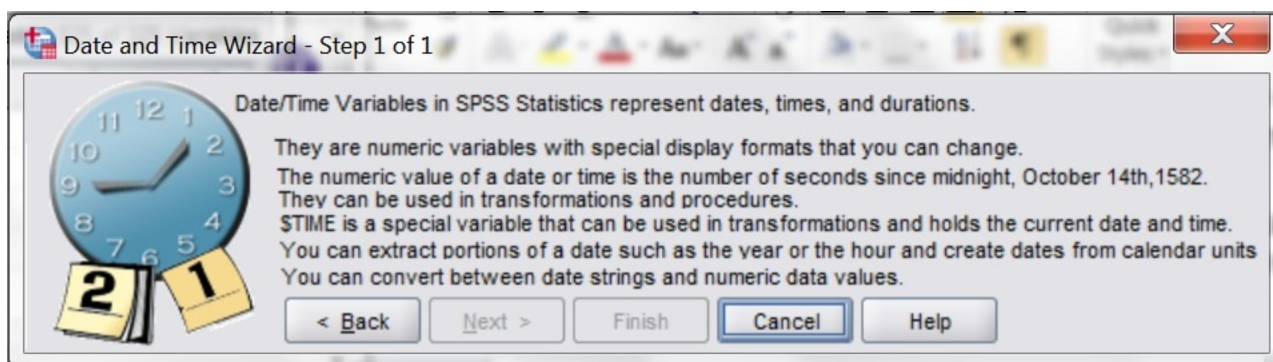
În programul SPSS, următoarea opțiune a meniului „Transform” este „Data and Time Wizard” care simplifică o serie de sarcini asociate cu variabilele care reprezintă timpul sau data calendaristică (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 7.44 Transform, Data and Time Wizard, în SPSS



Prima opțiune pentru „Time Wizard” este „Learn how dates and times are represented”. Dând comanda next vom fi conduși la o fereastră ce oferă o scurtă trecere în revistă a variabilelor dată / timp din SPSS Statistics. Dând click pe butonul „Help”, acesta oferă și un link către informații mai detaliate.

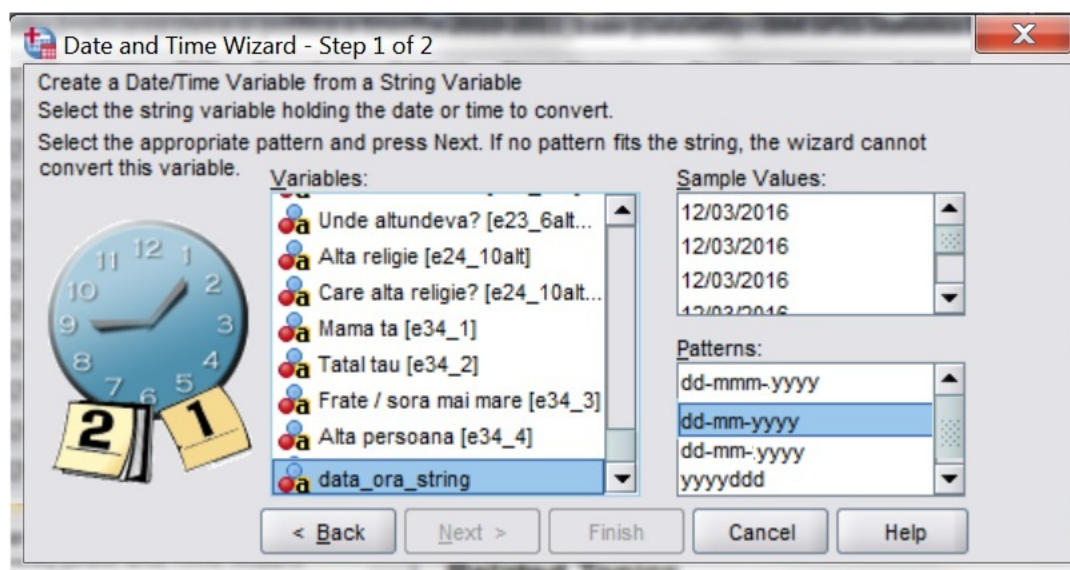
Graficul 7.45 Transform, Data and Time Wizard, pasul 1, în SPSS



„Create a date/time variable from a string containing a date or time” este a doua variantă pentru „Time Wizard” și permite crearea unei variabile de tipul dată / oră dintr-o variabilă string.

De exemplu, dacă există o variabilă alfanumerică care reprezintă datele în formatul lună/zi/an (LL/ZZ/AAAA), se poate crea o variabilă dată/oră din aceasta. În exemplul din graficul 7.46 am folosit o variabilă „string” având formatul lună/zi/an. În partea din dreapta a graficului 7.46 sunt listate variabilele string dintre care am selectat-o pe cea denumită „data_ora_string”. În dreapta sus au apărut mostre dintre valorile variabilei iar jos, posibile modele de reprezentare a datei și timpului (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Graficul 7.46 Transform, Data and Time Wizard, pasul 2, în SPSS

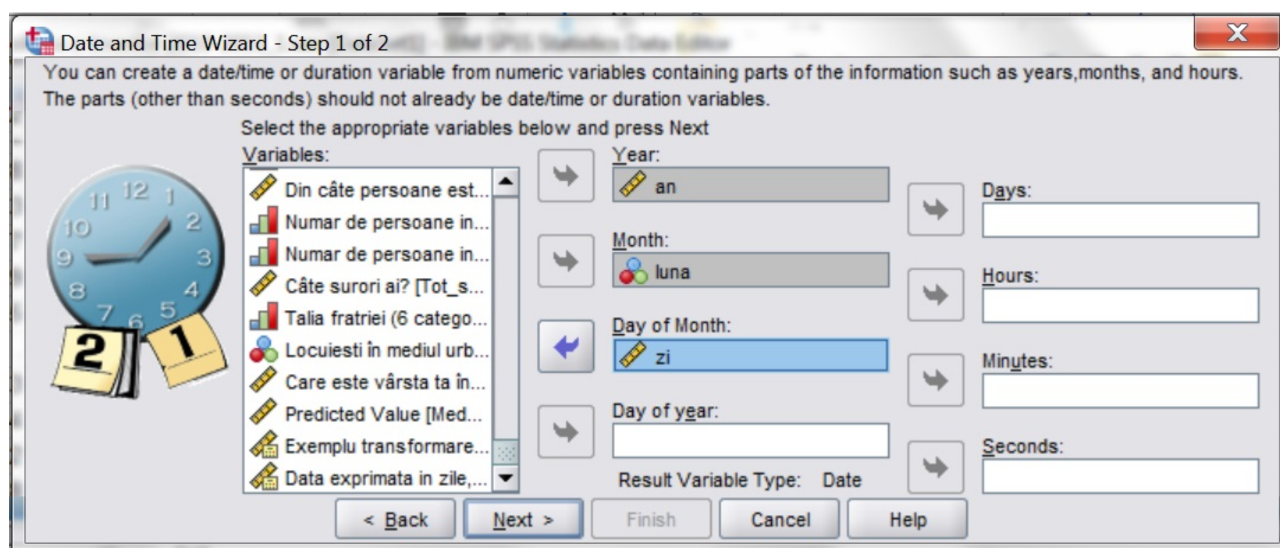


După ce s-a ales variabila alfanumerică ce va fi transformată în dată / timp precum și modelul de reprezentare a datei și timpului, se va da comanda next și se va obține o nouă fereastră în care se va specifica:

- Numele noii variabile de tip dată / timp care va fi construită
- Eticheta variabilei
- Dacă se va construi imediată noua variabilă sau se vor salva comenzile SPSS prin care se obține variabila într-o fereastră tip sintaxă (IBM Knowledge Center, Home, SPSS Statistics 24.0.0)

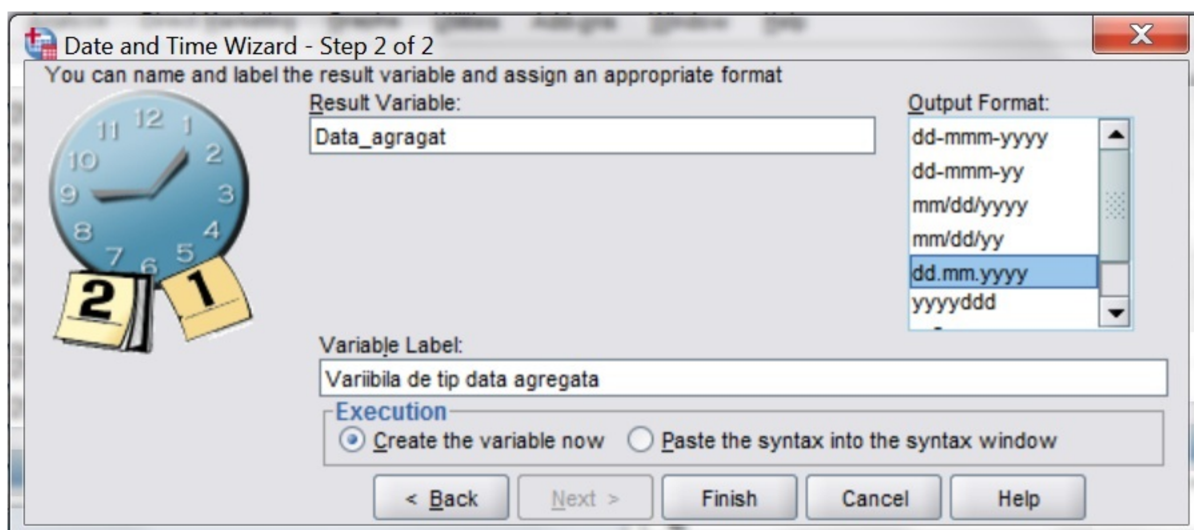
A treia variantă a opțiunii „Transform", „Data and Time Wizard" este „Create a date/time variable from variables holding parts of dates or times" și permite construirea variabilelor dată / oră dintr-un set de variabile existente. Dacă vom da click pe butonul „Next" vom obține fereastra din graficul 7.47. De exemplu, avem o variabilă care exprimă luna (ca număr întreg), a doua care desemnează ziua și o a treia care reprezintă anul. Putem combina cele trei variabile într-o singură variabilă de tipul dată / oră.

Graficul 7.47 Transform, Data and Time Wizard, pasul 3, în SPSS



În Graficul 7.47 am introdus variabilele cantitative „an", „luna" și „zi" în căsuțele corespunzătoare din fereastra de dialog. Dând comanda „next" vom obține o nouă fereastră în care putem specifica: numele variabilei de tip dată pe care o vom construi, eticheta precum și forma de reprezentare a acesteia.

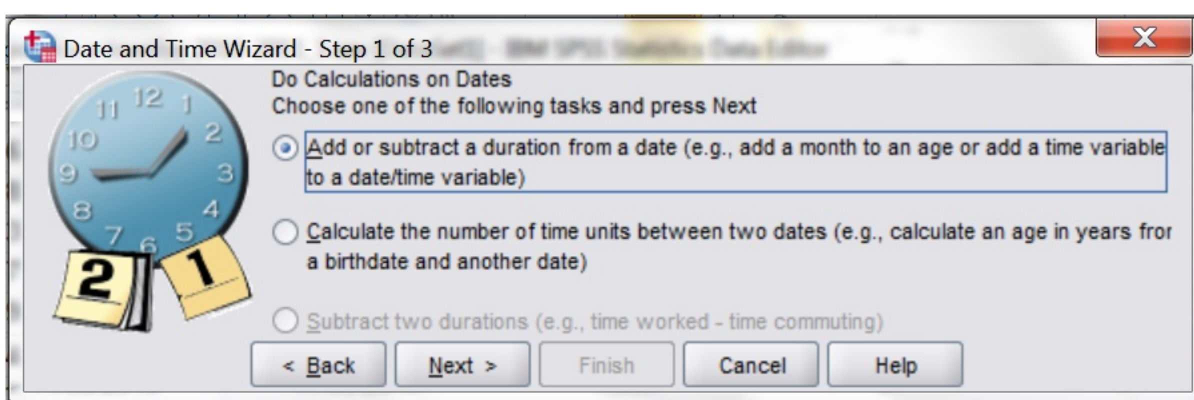
Graficul 7.48 Transform, Data and Time Wizard, pasul 4, în SPSS



În graficul 7.48 numele variabilei nou-constrită este „Data_agragat", eticheta acesteia e „Variibila de tip data agregata" iar forma de reprezentare a variabilei de tip dată e „dd.mm.yyyy (zizi.lunaluna.anananan).

A cincea posibilitate a opțiunii „Transform", „Data and Time Wizard" este „Calculate with dates and times" care oferă posibilitatea adăugării sau a scăderii unor valori din variabilele de tip dată / timp.

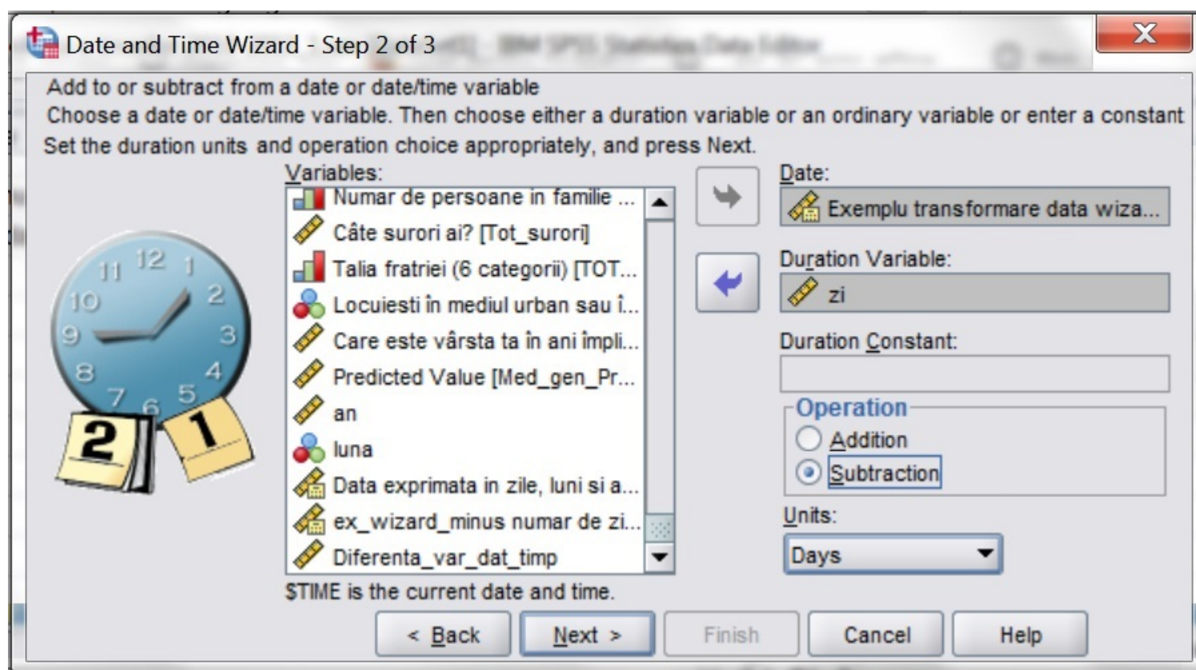
Graficul 7.49 Transform, Data and Time Wizard, pasul 5, în SPSS



După cum se vede în graficul 7.49, se poate alege varianta adunării sau scăderii unei durate dintr-o variabilă de tipul dată/timp sau se poate calcula numărul de unități de timp dintre două date (de exemplu, calcularea vârstei în ani între data nașterii și o altă dată de referință). Am ales să scădem o variabilă care exprima numărul de zile - „zi"- (introdusă în căsuța „duration variable) din variabila dată/timp construiă anterior Exemplu_data_wizard cu extensia „exemplu transformare data wizard (întrodusă în căsuța „Date" (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

În graficul 7.50 se vede că am ales, în căsuța „Operation”, varianta scădere (Substraction) iar pentru măsurarea timpului am ales zile („Units: Days”).

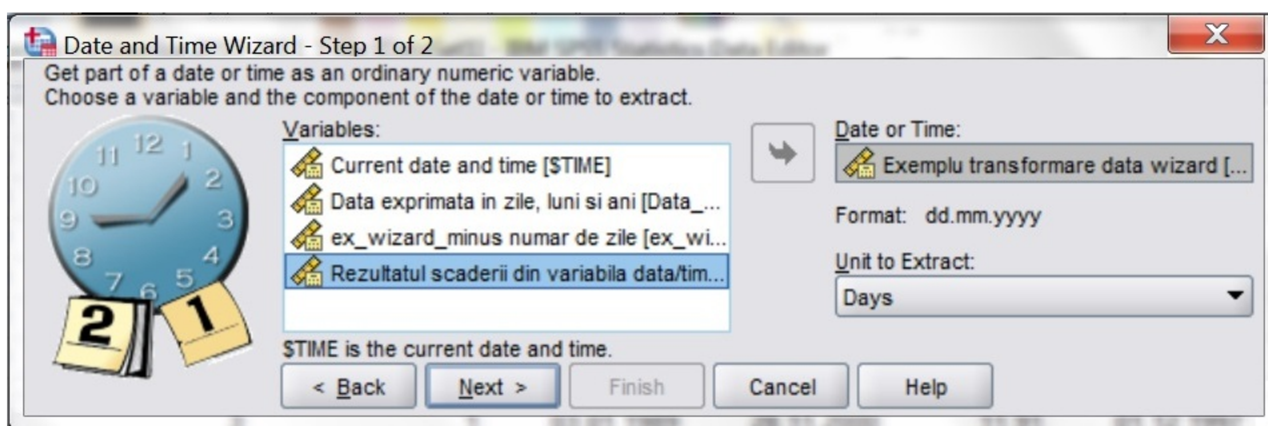
Graficul 7.50 Transform, Data and Time Wizard, pasul 6, în SPSS



Acționând butonul „Next” se va deschide o nouă fereastră în care vom trece numele variabilei ce urmează a fi construită și eticheta acesteia. Tot aici putem opta între a crea imediat acea vaiabilă și a salva comenzile SPSS proprii procedurii într-o fereastră sintaxă (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

A șasea variantă a opțiunii „Transform”, „Data and Time Wizard” este „Extract a part of a date or time variable” care oferă posibilitatea extragerii unei dimensiuni dintr-o variabilă dată/oră. De exemplu poate fi extrasă ziua dintr-o variabilă dată / oră care are forma zizi.lunaluna.anananan.

Graficul 7.51 Transform, Data and Time Wizard, pasul 7, în SPSS



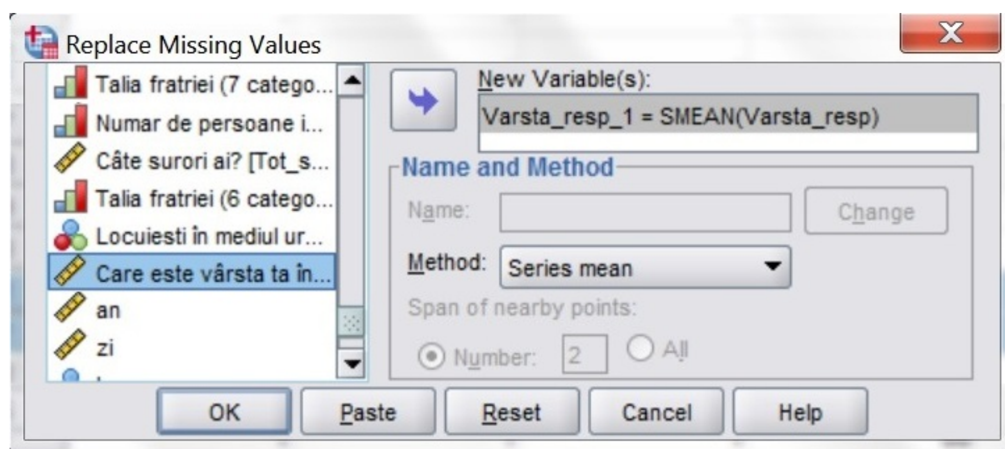
Dând click pe butonul next vom putea atribui un nume variabilei ce urmează a fi calculată și o etichetă. Și aici putem opta între crearea imediată a variabilei și salvarea comenzilor SPSS într-o fereastră tip sintaxă (IBM SPSS Statistics 20 Core System, User’s Guide, 2011).

A șaptea și ultima posibilitate oferită de opțiunii „Transform", „Data and Time Wizard" este „Assign periodicity to a dataset" care va lansa fereastra de dialog „Definire date", a meniului „Data", utilizată pentru a crea variabile de tipul dată / oră. Nu vom relua explicațiile deoarece pot fi regăsite în capitolul dedicat meniului „Data", „Define Dates", Graficul 6.20.

7.7 Opțiunile TRANSFORM, Create Time Series, Replace Missing Values în SPSS

Opțiunea „Create Time Series", a meniului „Transform", permite crearea unor noi variabile bazate pe caracteristicile variabilelor numerice de tipul serie de timp existente. Variabilele nou-create sunt utile în multe proceduri de analiză a seriilor de timp. Numele implicite ale noilor variabile sunt primele 6 caractere ale caracteristicii utilizată pentru a le crea, urmate de un număr ordinal. De exemplu, pentru „Pretul_lunar_apa", noua denumire a variabilei ar fi Pretul_1. Noile variabile păstrează etichetele definite pentru variabilele inițiale (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

Graficul 7.52 Transform, Create Time Series în SPSS



Există funcții disponibile pentru crearea seriilor de timp („Function") din care am ales sumarea cumulativă a valorilor („Cumulative sum"). După ce acționăm butonul „Ok", noua variabilă va fi construită, conform procedurii.

Variabila nou creată va oferi, pentru fiecare lună a perioadei definite, suma costurilor începând cu momentul inițial, definit de variabila serie de timp (IBM Knowledge Center, Home, SPSS Statistics 24.0.0).

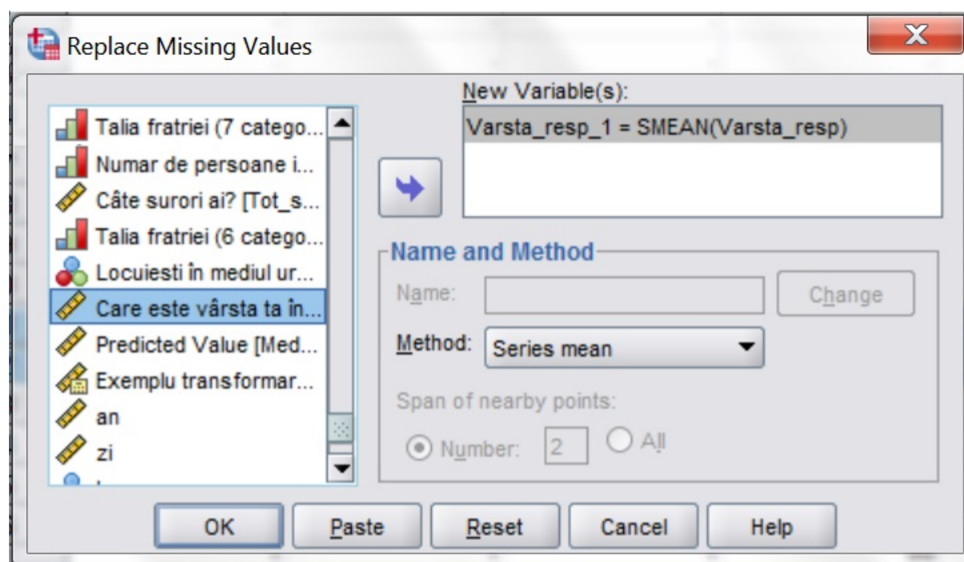
Următoarea opțiune a meniului „Transform" pe care o vom prezenta-o este „Replace missing values".

Procedura are rolul de a crea noi variabile pornind de la cele existente în baza de date activă, utilizând o anumită metodă de estimare a datelor lipsă. În graficul 7.53 este ilustrată fereastra de dialog care permite înlocuirea datelor lipsă folosind meniul „Transform". În stânga sunt listate variabilele din baza de date iar în dreapta, variabilele pentru care se va aplica procedura de înlocuire (selectate din stânga și transpuse în dreapta). Dacă se selecează mai multe variabile, pentru fiecare în parte va trebui indicată metoda de înlocuire a datelor lipsă (iar numele va fi atribuit automat: cel initial la care se adaugă „_" (liniuța din partea de jos) și valori numerice (IBM SPSS Statistics 20 Core System, User's Guide, 2011).

Variantele de înlocuire a datelor lipsă în fereastra „Replace missing values sunt următoarele:

- Series of means: se înlocuiesc datele lipsă cu media pentru întreaga serie
- Mean of nearby points: sunt înlocuite datele lipsă cu media valorilor valide din jur. „Span of nearby points" reprezintă numărul de valori valide deasupra și sub data care lipsește utilizate pentru a calcula media.
- Median of nearby points: se înlocuiesc datele lipsă cu valoarea mediană a valorilor valide din jur. „Span of nearby points" semnifică numărul de valori valide deasupra și dedesubtul datei lipsă luate în considerare pentru a calcula valoarea mediană
- Linear interpolation: sunt Înlocuite datele lipsă folosind o interpolare liniară. Ultima valoare validă anterioară datei lipsă lipsă și prima valoare validă după cea lipsă sunt utilizate pentru interpolare. Dacă prima sau ultima valoare din serie lipsesc, data lipsă nu va fi înlocuită.
- Linear trend at point: sunt înlocuite datele lipsă cu tendința liniară a seriei de date pentru acel punct. Se realizează o analiză de regresie în care variabila independentă ia valori de la 1 la N (totalul cazurilor). Valorile lipsă sunt înlocuite cu valorile lor estimate prin ecuația de regresie.

Graficul 7.53 Transform, Replace Missing Values în SPSS



Notăm faptul că pentru metodele de înlocure a datelor lipsă pentru care se specifică numărul de valori anterioare si succesive care sunt luate în calcul („Span of nearby points"), numelui variabilei generat automat i se va adăuga la sfârșit și această valoare.



ISBN: 978-606-37-0428-4
ISBN: 978-606-37-0429-1